390FinalPaper

Jillian Myler

2024-04-30

Introduction

Access to healthcare is vital. When a loved one is unwell or injured, our instinct is to support them. Similarly, when our own health hinders our daily tasks, seeking treatment becomes necessary to improve our quality of life. Health influences the daily routines of everyone and especially impacts those facing injury or chronic illness. Therefore, access to healthcare is paramount in ensuring the well-being of individuals and communities. Further, the access to healthcare should not be influenced by an individual's socioeconomic status. To give a concrete example of perpetuated socioeconomic disparity even within a publicly funded health care system (England) it is found that there is a 35% delta (43 days) between the most and least advantaged population groups in those seeking treatment for non-emergency coronary revascularization procedures [1].

The paper chosen for this project examines disparities in wait times across socioeconomic groups for patients that are seeking treatment for non-emergency coronary revascularization procedures. Wait times for non-emergency situations are strong indicators of the barriers to access that still exist within a healthcare system that removes price as a factor. Thus, it is crucial to determine if there is a significant difference in wait times across socioeconomic groups, and what may be the root cause of such a difference in order to inform future policy decisions towards equitable access to healthcare. Within their study, the authors of this paper also take into consideration that the patient may be making a choice that directly contributes to wait time differences. For example, when an individual chooses to forgo the hospital that is nearest to them and instead go to the next nearest hospital in an attempt to get either better service or a shorter wait time. Thus, the authors develop a metric that is meant to measure the differential distance between the closest and second closest provider. This distance metric is a strong predictor of the probability that the individual chooses to bypass the nearest hospital. The quantified probability of a patient's choice to go to one hospital versus the other to seek treatment is meant to try and encapsulate how patient choice factors into the wait time differences; therefore, leaving the remaining differences in wait time to be explained by bias by the provider or disproportionate access existing for those of higher SES.

[1] Moscelli G, Siciliani L, Gutacker N, Cookson R. Socioeconomic inequality of access to healthcare: Does choice explain the gradient? J Health Econ. 2018 Jan;57:290-314. doi: 10.1016/j.jhealeco.2017.06.005. Epub 2017 Jun 23. PMID: 28935158.

While the paper I selected aimed to quantify the amount of the difference in wait times between those of higher SES and lower SES due to choice in a publicly funded healthcare system for non-emergency cases, it is also relevant to consider wait times across socioeconomic statuses in more urgent cases where choice is already ruled out as a factor. The aim of this paper will be to discuss thoroughly disparities in access to healthcare across socioeconomic groups through the frameworks of statistics, economics and philosophy.

Methods

The main research question that aims to be answered in the paper of focus is: Do observed inequalities of access (for two treatments for a severe and costly disease) by socioeconomic status result from patients from different socioeconomic status making different choices?The authors write "Failure to account for patients' choices may lead to self selection bias in estimates of the direct socioeconomic waiting time gradient."(Moscelli et.al, 2018) This statement is the motivation behind the paper itself; therefore it is highly relevant that it is understood. Essentially, it is possible to see a sample in which people of lower SES have longer waiting times because they chose alternatives that entailed inherently longer wait times. Thus, they self-selected into those longer wait times. Resultantly, if those decisions are not accounted for, then it is possible to erroneously conclude that the correlation is a direct effect of SES on the waiting times.

To give context to this study, the National Health Service (NHS) of England experienced a period of accelerated expenditure growth because there was a perception that the NHS was underfunded. Thus, in a 5 year period starting in 2003, there was a 50% increase in budget allocation which then led to a 33% increase in treatment capacity across England. It follows that a massive increase in budgetary and treatment capacity would lead to massive efforts toward reforms in healthcare practices and policies. One policy that was subsequently introduced as a result of increased capacity was allowing patients to choose which hospital they would like to receive treatment at [2]. Hence,with a specific new policy related to choice, and knowledge of potential disparities already existing in wait times across socioeconomic status, it was an optimal time to see how much the new policy allowing choice seemed to contribute to such disparities in wait time.

The data used in this paper was for the 9 financial years (April to March) from 2002-2003 to 2010-2011 and the sample includes all elective patients for either CABG or PCI surgeries. The data was sourced from the Hospital Episode Statistics (HES) database which is a database containing all hospital admissions statistics in public hospitals in England. The purpose of the HES is to allow healthcare analysis, such as the analysis done in this paper, by national bodies and regulators including Public Health England, NHS Improvement, Integrated Care Boards and other groups with the intent of holding the public hospitals to a high standard of care. Initial data cleaning involved removing duplicates of patient records or incomplete patient records. Next, in order to begin to build indicators of a patient's socioeconomic status, the researchers used the Economic Deprivation Index which offers a proportion of individuals in the working age (18-59) in a given Lower Super Output

[2] Department of Health. The NHS Plan: a plan for investment: a plan for reform. Stationery Office, 2000.

Area (LSOA,roughly equivalent to a county in a U.S. State) who are living in low income households, or filing for social security benefits indicative of being unemployed. Using the EDI, the researchers generated dummy variables based on the quintiles of income deprivation of each LSOA each year. Next, in order to estimate the distances each patient would have to travel to each hospital, the authors chose to calculate the linear distance from the centroid of the LSOA that the patient lived in to the nearest two hospitals (by geographic coordinate), then took the difference between the two distances to be used in the selection equation. Further, the authors took an average of the distances of the three nearest hospitals offering the two different treatments (PCI and CABG) to the address of the patient. This metric is meant to allow for there to be a second level of choice via a Roy Model for Switched Regression.

Moving from data acquisition to actual modeling, the authors first employ a linear regression model with hospital fixed effects, estimated separately for each financial year and revascularization procedure. The model is defined as follows: $ln(W_ij) = h_j + \beta_1 y_ij + \beta_2 s_ij + \beta_3 x_ij + \epsilon_ij$

Here, W_ij represents the waiting time of patient i in hospital j, and y_ij denotes the vector of dummy variables representing socioeconomic status (SES) derived from the EDI as described above. SES is categorized into quintiles, with the highest quintile serving as the reference category. The coefficients β_1 capture the impact of SES on waiting times, with positive values indicating income-related inequalities favoring the affluent.

The vector s_ij includes severity controls such as age, gender, number of secondary diagnoses, previous hospital admissions, and Charlson comorbidities. These severity proxies are important to include because they are able to capture the patient's health status, which is unknown to the econometrician, but known to the patient or their doctor. Thus, adding these proxies controls for legitimate differences in waiting times that may emerge from patients with lower health having priority in the waitlist. Additionally, the vector x_ij incorporates non-severity variables like the month of admission.

The hospital fixed effects h_j account for differences in waiting times among hospitals, capturing unobserved demand and supply factors such as infrastructure, staffing, and clinical quality. Consequently, β_1 should be interpreted as waiting time inequalities within a hospital, rather than across hospitals. The analysis focuses on socioeconomic inequalities within hospitals, as controlling for hospital fixed effects reveals minimal changes in waiting time gradients across hospitals in England. This indicates that waiting time inequalities predominantly arise within hospitals. It is important to note that using a hospital fixed effect also implies that the individuals of lower socioeconomic status are not systematically living in locations where their nearest hospitals have longer wait times.

Since the authors are trying to define the influence of choice on wait times, they must also factor in how the patient choice between the two treatments may impact wait times. The Roy model regression, also known as the switching regression model, serves to address self-selection bias in observational data, particularly in situations where individuals make choices that affect the outcome being studied. Within this study, the choice is two-fold: which treatment option (CABG or PCI) and whether or not to bypass the nearest hospital to a given patient. Thus, a Roy model regression is used within this paper to fully understand the effect of patient choice. The particular models used within their paper are described below.To model the choice of hospital, a selection equation is introduced to capture the decision to bypass the closest hospital. n_{ij} is defined as a dummy variable equal to 1 if the patient bypasses the closest hospital and 0 otherwise. The Roy model for hospital choice is then formulated as:

$$n_{ij} = I(\lambda_0 z_{ij} + \lambda_1 y_{ij} + \lambda_2 s_{ij} + \lambda_3 x_{ij} + v_{ij} > 0), \quad n_{ij} = \{0, 1\}$$

where w_{ij1} and w_{ij0} represent the observed log waiting times for patients selecting respectively into the non-closest or closest hospital, and w^*_{ijn} is the latent waiting time outcome for every patient before self-selecting into a given hospital.

$$w^*_{ijn} = \begin{cases} w_{ij1} = h_j + \tilde{\beta}_{1,1} y_{ij1} + \tilde{\beta}_{2,1} s_{ij1} + \tilde{\beta}_{3,1} x_{ij1} + \varepsilon_{ij1}, & \text{if } n_{ij} = 1 \\ w_{ij0} = h_j + \tilde{\beta}_{1,0} y_{ij0} + \tilde{\beta}_{2,0} s_{ij0} + \tilde{\beta}_{3,0} x_{ij0} + \varepsilon_{ij0}, & \text{if } n_{ij} = 0 \end{cases}$$

The equation $n_{ij} = I(\lambda_0 z_{ij} + \lambda_1 y_{ij} + \lambda_2 s_{ij} + \lambda_3 x_{ij} + \epsilon_{ij} > 0)$ defines the patient's decision ($n_{ij} = 0$ or $n_{ij} = 1$) to bypass or not the closest hospital. The expected waiting times conditional on hospital choice are given by:

- $E(w_{ij1}|h_j, y_{ij1}, s_{ij1}, x_{ij1}, \hat{p}_z) = h_j + \beta_{1,1} y_{ij1} + \beta_{2,1} s_{ij1} + \beta_{3,1} x_{ij1} + \epsilon_{ij1}$ if $n_{ij} = 1$
- $E(w_{ij0}|h_j, y_{ij0}, s_{ij0}, x_{ij0}, \hat{p}_z) = h_j + \beta_{1,0} y_{ij0} + \beta_{2,0} s_{ij0} + \beta_{3,1} x_{ij0} + \epsilon_{ij0}$ if $n_{ij} = 0$

The unobserved error terms $v, \epsilon_0, \epsilon_1$ follow a degenerate trivariate Gaussian distribution with mean zero and covariance matrix n. The covariance between waiting times when choosing different hospitals is not defined, as only one outcome (either w_{ij1} or w_{ij0}) can be observed at any given time. Patients are assumed to self-select into hospitals based on factors like severity, co-morbidities, age, and income deprivation, seeking the highest latent utility. The model includes equations for observed and latent waiting times, controlling for hospital characteristics through residual terms v_{ij} in the waiting time equations.

By employing a combination of regression modeling and the Roy model regression, the research aims to fully embody wait time discrepancies based on difference in choice to forgo the closer hospital or not. Applying these models, the authors found that patients residing in the most deprived fifths of small areas experience a 35% longer wait time compared to those in the least deprived fifths. In 2002, the wait time for coronary artery bypass grafting (CABG) was 53% longer for these patients, while for percutaneous coronary intervention (PCI), it was 35% longer. These numbers decreased to 9.5% and 15% respectively by 2011. Further, the Roy regression models allowed for the authors to determine that waiting time inequalities are not primarily due to choice of hospital or type of treatment for the life-threatening condition investigated. Only up to 12% of the overall waiting time gradient (35%) was found to be due to patient choice in 2002. Therefore, it is clear that the difference in wait time across socioeconomic status is not accounted for by patient choice and instead must be due to bias by providers or advantages inherent to those of higher SES.

Extension

The HES has a large amount of publicly available data, however, it is important to note that the data that allows for the calculations of distance of an individual to the nearest or second nearest hospital or to determine the severity level of the individual's health is not accessible without higher levels of permission. In looking to re-create this analysis, or look into something new to be added to the model within the same data, I found many publicly accessible CSVs related to dementia or autism or even emergency room admissions, but address or income level at the patient level were certainly always left out, making a replication infeasible. Instead, I decided to look into a few of the factors placed into controls in the original regression model in healthcare data that I could access.

My first ambition was to create my own data set based on U.S. City population data insofar as being able to easily access the population of a city as well as measure density[3]. In this effort, I also found data on average wait times for non-emergency procedures across 5 different specialties in 15 of the most populated cities in the U.S. from the Survey of Physician Appointment Wait Times (2009,2014,2017,2022)[4]. Finally, the Pew Research Center has compiled a resource of open data that has compiled all of the hospitals in the U.S. (as of 2022) and allows for a filter by city name and state to find a count of hospitals[5]. The aim in this effort was to compile a dataset to look at whether city density or count of hospitals in comparison to density had an impact on wait time in a country that does not have a predominantly publicly funded healthcare system. A final addition to the dataset was the percent of the population in each city that had an income less than half of the poverty threshold, half to However, the dataset was based on averaged values given in summary tables. The result was only an 15 rows (one for each city) and 11 columns.

The reality of a data set that has a sample size of 15 is that no meaningful regression or modeling can really be performed because the sample size is too small. I attempted a multiple linear regression model just to see what the sign on the estimated parameters were. In other words, just to see if it seemed density (people per square kilometer), amount of hospitals, or income level relative to poverty threshold had increasing had an estimated negative or positive impact on expected wait time. No models garnered an r-squared value of more than 0.14, or had a single variable significant at even the p=0.1 level. I considered bootstrapping with the aim of not increasing the sample size, but resampling without replacement, generating a linear model and then averaging the parameter estimates after each round to fine tune the parameter estimates more. However, even after 1000 rounds of bootstrapping, the aggregate parameter estimates were approximately zero. In other words, the small amount of data in the original data set I compiled made modeling or regression a fruitless action.

Next, after consulting kaggle, a new data set was found [6]. This data set is called **Hospital Patient data**. This data represents patients admitted to the Emergency room in a U.S. hospital. The variables include *Date*, *Medication Revenue*, *Lab Cost*, *Consultation Revenue*, *Doctor Type*, *Financial Class*, *Patient Type*, *Entry Time*, *Post Consultation Time*,*Completion Time* and *Patient ID*. At this point, I acknowledge no novel modeling would occur, nor

[3]"United States Cities Database." Simplemaps, simplemaps.com/data/us-cities. Accessed 3 May 2024.

[4]Benson, Craig. "Poverty in States and Metropolitan Areas: 2022." American Community Survey Briefs, Dec. 2023, https://doi.org/https://www.census.gov/content/dam/Census/library/publications/2023/acs/acsbr-016.pdf.

[5]Arcgis.Com, www.arcgis.com/home/item.html?id=75079bdea94743bcaca7b6e833692639#data. Accessed 3 May 2024.

[6]Abdulqader_Asiirii. "Hospital Patient Data." Kaggle, 14 Apr. 2023, www.kaggle.com/datasets/abdulqaderasiirii/hospital-patient-data?resource=download.

would there be a proof necessarily of their methods. Instead, the aim with this new data was to understand what wait time looks like in the U.S. based on the *Financial Class* variable. Further, one of the variables within the paper's model (x_{ij}) that captures non-severity factors such as day/month of admission. This particular data set also allows me to look at difference in wait times based on day of week. Since the paper identified a clear disparity remaining in wait time based on socioeconomic status, this data set offers an opportunity to see if there is a similar situation in the United States, but also allows for looking more into the control variable x_{ij}.

Summary Table

financial_class	mean_waiting_time	number_of_patients
CORPORATE	46	6915
HMO	46	3738
INSURANCE	44	9931
MEDICARE	58	293
PRIVATE	40	9121

Summary Table

weekday	mean_waiting_time	number_of_patients
Friday	42	4923
Monday	49	6982
Saturday	43	3010
Sunday	33	2549
Thursday	42	2673
Tuesday	42	5690
Wednesday	47	4171

Above, there are two summary tables. The first, shows the average wait time of individuals based on their *financial class*. In this case, *financial class* describes the type of insurance that the patient has. For reference, in the United States, the health care system is largely privatized and is largely paid for through third party insurance companies. Thus, *Corporate* implies that the individual will have most of their medical expenses paid by their corporate insurance policy, a policy provided for the company that they work for, *HMO* (Health Maintenance Organization) is a limited health insurance plan that applies only with the doctors or providers that are a part of the HMO. Further, *Insurance* differentiated from *Private* indicates that *Private* is meant to capture only those insurance policies that are completely privatized (more expensive generally) while the more broad *Insurance* embodies other group based insurances. Finally, it is important to contextualize *Medicare* as the United States' welfare public insurance program for individuals at a certain level of poverty. Thus, this table shows that in this Kaggle data set, there is an 18 minute difference in the average wait time for individuals with private coverage, generally representative of the individuals in the highest socioeconomic standing, and those who are covered by Medicare. The waiting times increase accordingly with expected income level based on insurance status; therefore, at least in this data set from the United States, the disparity in wait time is consistent in both continents/societies. Further, looking at the next summary table, there is variance in wait time based on day of the week. The largest delta in wait time being admitted on a Monday versus being admitted on a Sunday. This finding just backs up the need for a variable accounting for non-severity variables such as x_{ij}. Finally, given that this data was from an emergency room setting, the patient choice between one hospital and another and between treatments that was explored within the paper is no longer a factor here.

Normative Concerns

Access to healthcare is not only an economic and societal issue, but also an ethical issue. With regards to normative considerations in the issue of accessibility to healthcare, perspectives from both Immanuel Kant and John Rawls can be applied. Immanuel Kant is a philosopher whose doctrines fall under deontology, the study of the nature of duty and obligation. One of his main doctrines in such duty ethics is the categorical imperative, which is two-fold. Firstly, the categorical imperative requires that individuals are treated as ends and not means. Second, the categorical imperative requires that the action taken is only justifiable such that one can universalize it. When observing a gradient in healthcare access across socioeconomic status, it is clear that individuals are being discriminated against or that within the current healthcare system (UK in the case of the paper) there are imperfect agents who are not viewing each patient as an end. Instead, there are other factors that are causing individuals of identical need to be prioritized over others. This extended wait period for those of lower SES is shown to exist in the U.S. as well, where medicine is privatized, allowing for greater potential external incentives facilitating such imperfect agency. Further, the quicker treatment times reflected in those of higher SES are not universalized. Thus, a gradient in wait times across SES groups directly violates the categorical imperative of what a healthcare system, according to Kant, is obligated to provide.

While Kant is known for his duty-based ethics, John Rawls writes from a lens of justice ethics. Rawls paints the concept of justice as fairness and defines a difference principle that I will apply to this case as well. The difference principle suggests that in situations where disparities exist, resources should be allocated to protect the most vulnerable. If we were to define disparity as difference in SES grouping, the outcomes of this paper directly violate this difference principle because those of lower socioeconomic status are experiencing worse accessibility to healthcare than those of higher SES. It is worth clarifying that in healthcare, it is not that one group should have access over another, but instead that if there is a disparity existing that is disproportionately affecting the less privileged, then the difference principle has been violated.

Finally, in the process of searching for data for this paper, differing levels of accessibility to data became apparent. In this class, there was a lot of discussion on the trade off between privacy and accuracy. I found the data source of my selected paper to be an interesting case study in a similar trade off. The Hospital Episode Statistics (HES) and the entire data section of the National Health Services Digital has such a large database of health care information. It is commendable to have such a database with the intent of making data available to policy makers or those who are intending to do research to better the efficiency and span of the healthcare system as a whole. There is so much available in the form of summary tables in any specialty you can think of. However, the NHS does their due diligence to protect individual patients as well because there is no trace of a patient's individual information without requesting specific access and defining what organization you are a part of, as well as what is intended to be done with the data. This system seems to be a commendable program because it intends to be open and transparent with the public, given that the public both funds and receives care from its services. Further, the data is collected to allow research to keep their healthcare system accountable. In other words, if the data is to be used to help better the system, under contract with those who released the data to use it just for that purpose, it seems that this demonstrates a consequentialist ideal usage of data because of the great societal benefit.

Healthcare data is reasonably protected because it contains sensitive information, which in a privatized healthcare system potentially offers more risk. It is important to also discuss how if the individual data was acquired and leaked to another group with an aim that is not to improve the healthcare system as a whole. For example, a tech company getting access to individual health records in a privatized healthcare system and being able to see a potential hire as more expensive because of their health history, this would allow for job market discrimination. Hence, this extensive healthcare database (HES/NHS Digital) demonstrates highly beneficial usages of data accessibility, but it's protections, and any other healthcare database's protections, must be strong or there is potential for harm in the labor market and beyond. In accordance with the harm principle, these sources are only commendable until they cause harm.

Conclusions

Access equity is a crucial goal in publicly-funded healthcare systems. Yet, disparities in access based on socioeconomic status may stem from variances in patient decision-making. The paper's analysis of non-emergency coronary revascularization procedures in the English National Health Service revealed significant differences in wait times among patients of different socioeconomic statuses. These variances amounted to a maximum of 35%, or 43 days, between the most and least deprived quintile groups. Employing selection models with differential distances as identification variables, we determined that only a maximum of 12% of these wait time disparities could be ascribed to patients' choices regarding hospitals and treatment types (such as heart bypass versus stent). Substantial residual inequality persisted even after accounting for patient choice, particularly benefiting patients in the least deprived quintile group with shorter waiting times.

Upon sourcing data for individual patients admitted to the Emergency room in a hospital in the United States, similar evidence of longer wait times for individuals of lower socioeconomic status was found. This data removes choice as a whole because the individuals are at the emergency room, indicating that the demand for care is high and that they were likely not considering wait time at all in seeking treatment. Further, the data from the United States uses insurance type in a privatized healthcare system, allowing for this disparity to potentially be more representative of systematic external monetary incentive. Finally, the data from the United States emergency room shows that there is certainly a necessity for the control variable that accounts for non-severity factors specific to the individual and hospital by showing that there is a difference in wait time based on day of week of admission.

Access to healthcare is indispensable, as it is our instinct to provide support to loved ones when they are unwell or injured. Similarly, when our own health issues impede our daily activities, seeking treatment becomes essential to enhance our quality of life. Given that health affects the daily lives of all individuals and has a heightened impact on those dealing with injury or chronic illness, ensuring access to healthcare is crucial for the well-being of both individuals and communities. Moreover, access to healthcare should remain unaffected by an individual's socioeconomic status.Furthermore, ensuring data accessibility for transparency and the advancement of the healthcare system as a whole is imperative. Without accountability and transparency, issues within the system can proliferate unchecked. However, it's equally crucial to monitor the usage and protection of data to prevent any perpetuation of systematic struggles for individuals with poor health, while also safeguarding sensitive information to maintain privacy.