Study Objective:
The present inquiry focuses on Medicare fee-for-service beneficiaries who suffered from kidney and urinary tract infections with major complications or comorbidity in the City of Baltimore, Maryland, in 2015. In particular, the goal is to discover whether hospital charges in these instances correlate to neighborhood median income and/or the degree of urbanization (or rurality) of the hospital locations.

Data and Method:
Three datasets formed the basis of this study. The large dataset (201,876 rows and 12 columns) containing the hospital charges relevant to our inquiry was obtained through the website of the Center for Medicare and Medicaid Services[1]. The median household income information for the specific zip codes that fit the study came from the U.S. Census Bureau. The RUCA file, which describes the degree of rurality and urbanization of a given zip code, came from the link from the University of North Dakota. All three files were first imported and processed by using the SAS Studio for data analytics. The data pertaining to the relevant disease code and location information were selected and extracted through using the SQL capabilities of SAS, resulting in a total of 9 observations for the 9 hospitals in Baltimore. The three files were then merged together by Provider_Zip_Code. There was a mistake in the original PUF file, which had assigned a non-existent zip code to Johns Hopkins Hospital on North Wolfe Street in Baltimore. During the data cleaning process, that was corrected so the data from the other two files could be utilized for our analysis correctly. Linear regression analysis was then performed on the final cleaned and merged dataset using the SAS Statistical Tasks and Utilities function.

Results:
Based on the results of the linear regression analysis, our fit plot with 95% confidence limits containing 9 observations indicates a negative correlation between the hospital charges and the median household income of the neighborhoods of the hospitals. In other words, the higher the median household income of a neighborhood, the lower the hospital charges were in the case of kidney and urinary tract infections for patients with major complications and comordibity. The coefficient of determinants (R squared) which indicates the proportion of variance in the dependent variable (hospital charges) which can be predicted from the independent variable (median household income) is 0.4228.[2] On the other hand, the rurality index of the locations of the hospitals in our study is uniform, with every value measuring 1.1. Based on the definition from the Rural Health Research Center, this value indicates a metropolitan area core.[3] Because of this uniform RUCA for all of our hospitals in the study, we could not answer the question of whether the metropolitan status of a hospital had an impact on its fees for treating patients.

Conclusions
Based on our statistical analysis, we conclude that in 2015 the hospitals in the poorer neighborhoods of Baltimore charged higher fees for treatment related to kidney and urinary tract infections combined with major complications and comorbidity, than their counterparts in more affluent neighborhoods. However, we do not know if all patients in the study were residents of the neighborhoods surrounding the hospitals. So, we cannot conclude from this data alone whether hospitals charged poorer patients a higher amount for the same services. Since we only analyzed data for one calendar year, we cannot conclude whether this is a situation unique to 2015 or one that has repeated itself. Furthermore, to study whether the metropolitan status of a hospital had an impact on its charges for services, we would need to expand our geographical region to include some rural counties near Baltimore. All in all, while the negative correlation found in our study is alarming, we need more data to make an informed recommendation for public policy changes.

---

[1] Inpatient Utilization and Payment Public Use File (PUF)

[2] Adjusted R Squared: 0.3404. Thus, about 34% to 42% of the variance in hospital charges can be explained by our model. This means that there were also other factors affecting our dependent variable.

[3] Primary flow within an Urbanized Area (UA) and secondary flow 30% through 49% to a larger UA, which could mean the tie to the Washington DC metropolitan area.