Variability of length of stay for Diabetic patients in the year 2019 and 2021 in the NY Dataset of hospital inpatient discharges (SPARCS)

Group Blue : (Niccolò Caselli, José Escarramán, Cristian Perrone)

ABSTRACT

INTRODUCTION: Diabetes is a condition characterised by high blood sugar levels resulting from the body's inability to produce or effectively use insulin. The pancreas plays a crucial role in diabetes as it produces insulin, a hormone that helps regulate blood sugar levels. We aim to describe the main factors that can affect excessive in-hospital length of stay of diabetic patients by focusing on differences among counties and social and clinical aspects of the patients. We also analysed how it changed between 2019 and 2021, before and after the outbreak of the Covid-19 pandemic. (Miao et al. 2022), (Katipoglu et al. 2022)

AIMS:

- What are the specific risk factors contributing to prolonged hospitalisation among diabetic patients?
- Is there excessive geographical variation of high length of stay between hospitals in the State of New York?
- How did the COVID-19 pandemic affect hospitalisation of diabetic patients?

METHODS: The length of stay variable (LOS) was dichotomised by using the upper quintile of 2019 distribution as cutoff, 1 for a LOS greater than 8 days, 0 otherwise, obtaining LOS8. Using the NY hospital inpatient discharge datasets from 2019, we conducted the analysis using Generalised Linear Models (GLMs) and Generalised Estimating Equation Models (GEEs) to identify potential risk factors associated with excessive length of stay. The likelihood ratio test and accuracy metrics were employed to compare models obtained through backward elimination. Direct and indirect standardisation methods were utilised to assess the hospitals' performances , and comparisons were made among hospitals in various areas. A funnel plot was used to focus on outlying observations. To assess if COVID-19 affected hospitalisation, we also considered the 2021 dataset, and in the model, we adjusted for the year difference between 2019 and 2021 as well.

RESULTS: A total of 44,423 diabetic patients for 2019 and 40,388 for 2021 were identified from the 2019 and 2021 NY SPARCS datasets. By focusing on 2019, undergoing amputation (OR=4.08 (3.81,4.36)), presenting major or extreme severity (OR=3.56 (3.32,3.81)), and being more than 70 years old (OR=1.86 (1.72,2.02)) appeared positively independently associated with LOS8. By fitting a logistic model on the 2019 dataset where the outcome is LOS8, being white, being male, severity, risk, age class, amputation, PCI and dialysis resulted statistically significant. There were no significant differences for what concerns excessive length due to diabetes-related complications. By fitting a GEE model with the same covariates, clustered for hospitals, we obtained similar results. In our analysis of LOS8 rates, we utilised indirect standardised rates, adjusting for age and sex. Furthermore, by adding the year difference between 2019 and 2021 in the previous model to test pre-post COVID-19 differences, this term resulted in statistical significance. Geographical analysis using standard rates revealed higher LOS8 rates in densely populated areas, particularly in southern New York counties like Long Island. Outlier hospitals, mainly in major population centers such as Kings, Queens, Manhattan, Suffolk, Bronx, and Nassau, may suggest potential disparities in healthcare provision.

CONCLUSIONS: The analysis identified significant predictors for prolonged hospital stays among diabetic patients, emphasising the importance of considering age, condition severity, and other factors in patient management. When comparing 2021 to 2019, there also seems to be an effect hinting at a correlation between Covid-19 and the length of stay of diabetic patients. Geographical analysis revealed disparities in hospital stays across New York counties.Future research should explore pandemic-related disruptions, including telemedicine's role, in improving diabetic patient care.

Introduction

Diabetes is a condition characterised by high blood sugar levels resulting from the body's inability to produce or effectively use insulin. The pancreas plays a crucial role in diabetes as it produces this hormone, which helps the regulation of blood sugar levels. Located behind the stomach, the pancreas contains specialised cells called islets of Langerhans, which release insulin into the bloodstream to manage glucose levels. When these cells malfunction, it often leads to diabetes. There are two main types of diabetes: type 1 and type 2. Type 1 diabetes occurs when the body's immune system attacks and destroys the insulin-producing cells in the pancreas. This results in a lack of insulin production, leading to high blood sugar levels. Type 2 diabetes, on the other hand, occurs when the body becomes resistant to the effects of insulin or doesn't produce enough insulin to maintain normal glucose levels. This can be caused by a combination of genetic and lifestyle factors, such as obesity, poor diet, lack of physical activity, and advancing age. Symptoms of diabetes can vary but may include increased thirst and hunger, frequent urination, fatigue, blurred vision, slow wound healing, and unexplained weight loss. Like pancreatic cancer, diabetes can often remain asymptomatic for long periods, earning it the moniker of a "silent disease." While the exact cause of diabetes is not fully understood, several risk factors have been identified. These include genetic predisposition, family history of diabetes, obesity, sedentary lifestyle, poor diet high in processed foods and sugars, and certain medical conditions such as high blood pressure and metabolic syndrome. Diabetes is a significant public health concern worldwide. According to the International Diabetes Federation, approximately 463 million adults were living with diabetes in 2019, with projections estimating this number to rise to 700 million by 2045 if current trends continue. Complications from diabetes can be severe and include cardiovascular disease, kidney failure, retinopathy, neuropathy, and lower limb amputation. Managing diabetes involves a combination of lifestyle changes, such as adopting a healthy diet, engaging in regular physical activity, monitoring blood sugar levels, and taking medications as prescribed. While there is currently no cure for diabetes, early detection and treatment can help individuals lead healthy and fulfilling lives while reducing the risk of complications. Ongoing research into the prevention and management of diabetes is essential to address this growing global health crisis and improve outcomes for those affected by the disease. Furthermore, the COVID-19 pandemic poses significant risks for individuals with diabetes, aggravating symptoms and complicating management. Factors such as compromised immunity and disruptions to healthcare services heighten these risks, emphasising the need for tailored support to safeguard the health of diabetic individuals during these challenging times.

Objectives

- To find out which are the specific risk factors contributing to prolonged hospitalisation among diabetic patients.
- To find out if there is a significant geographical variation of high length of stay across hospitals in the State of New York.
- To assess if COVID-19 pandemics affected hospitalisation of diabetic patients, considering the 2019 and 2021 NY SPARCS datasets

Materials & Methods

The data analysed in this study comes from the 2019 and 2021 New York Dataset of hospital inpatient discharges provided by the Statewide Planning and Research Cooperative System (SPARCS). The 2019 dataset contains information about N=2,339,462 patients admitted in NY area hospitals in 2019, while the 2021 dataset contains information about N=2,090,321 individuals admitted in 2021. Each hospitalisation is treated as an individual entry in the database, and is coded with 478 diagnoses for both years, and 316 and 322 procedures respectively for 2019 and 2021. Both datasets include 33 variables, which provide details about demographic information, clinical characteristics of the patients and their hospital outcomes. In order to pursue the aim of this analysis, our group considered a population of interest of N=44,423 individuals obtained by the 2019 dataset, and N=40,388 by the 2021 dataset, who were identified selecting only patients with “Diabetes mellitus with complications” and “Diabetes mellitus without complications” as diagnosis description. We first modified the length of stay variable to make it dichotomous, by selecting the last quintile of its 2019 distribution as cutoff. From the analysis we obtained a cutoff of 8 days of in-hospital length of stay, and so the variable LOS8 was defined. The number of diabetic patients whose in-hospital length of stay is greater than 8 days is N=8501 in 2019 and N=7808 in 2021 NY SPARCS dataset. Our group also created a series of dichotomous variables that could be interesting to analyse in the multivariate risk analysis:

- white: 1 if patient white, 0 otherwise
- d_risky, d_severe: dummy variables for major or extreme risk/severity
- d_admtype: 1 if admission type is an emergency, urgency or trauma, 0 otherwise
- PCI: 1 if the subject underwent percutaneous coronary interventions, 0 otherwise
- dialysis: 1 if the patient is in dialysis, 0 otherwise
- ampt: 1 if the patient experienced amputation due to neuropathy’s complications, 0 otherwise
- compl: 1 if the patient has diabetes-related complications, 0 otherwise
- recl_age: recodification of cl_age where the first 3 age categories were collapsed obtaining 0-49,50-69,70+

Exploratory Analysis: Initial examination of the study population through univariate and bivariate descriptive analyses to understand patient characteristics in 2019 and 2021.

Multivariate Logistic Regression Models: Following the main descriptive statistics, our group conducted multivariate logistic regression models to identify independent risk factors for prolonged hospitalisation in diabetic patients. We compared a binomial regression model within the family of Generalised Linear Models (GLM) to a Generalised Estimating Equations Model (GEE), both fitted using the 2019 NY SPARCS dataset. The GEE model, an extension of the GLM, incorporates information about the correlation structure within the data. Considering our scenario, accounting for correlation within hospitals could be interesting. Among various correlation matrix structures, we employed the exchangeable structure, which assumes uniform correlation across all hospital pairs. The choice of a binomial model stemmed from our outcome being dichotomous (yes or no). In the models, a positive coefficient suggests a potential risk factor, whereas a negative coefficient implies a possible protective effect.

Variable Selection: Once estimated two complete models, in which we adjusted for all the possible covariates explaining excessive length of stay, through the technique of Backward Elimination, we started taking out one variable at a time from the model, beginning with the covariate which showed the lowest degree of significance. We repeated this operation until all the predictors included in the model resulted statistically significant, with a p value lower than 0.05. At this point, we compared the initial complete model with the reduced one using the log Likelihood Ratio Test (LRT), to evaluate which model to prefer. LRT provides the possibility to compare different nested models, taking into consideration both the complexity of the model and how it fits well the data:

$$LRT = 2\ln(\mathcal{L})FULL - 2\ln(\mathcal{L})REDUCED$$

where $\mathcal{L}$ are the estimated likelihood functions of the complete and reduced models.

Standardisation Methods: Direct and indirect standardisation methods were applied to investigate the length of stay trends among different hospitals. Direct standardisation involved adjusting for age, gender, and county population, while indirect standardisation adjusted for age and gender. With Direct standardisation comparisons between different hospitals could be made, while for indirect standardisation just comparisons with the average value were possible. These methods facilitated proper comparisons among different facilities and hospitals within the same area by using a standardised population for comparison.

To assess COVID-19 differences: To assess differences in prolonged length of stay associated with COVID-19, we examined data from both the 2019 and 2021 SPARCS NY datasets. We created a dichotomous variable which takes into account whether a diabetic patient was discharged in 2019 or in 2021. We added this covariate to the existing GLM and GEE models to test its significance in both models. If the variation of prolonged length of stay between 2019 and 2021 results are statistically significant, then this would mean that the COVID-19 pandemic had a relevant effect on the phenomenon of interest.

Results

Exploratory Analysis: We start by presenting an exploratory analysis of the dataset used in this study to provide insights into the demographic and hospital characteristics of the study population. We will consider the 2 datasets separately here, as this gives a more precise look at the populations that are considered in the following analyses. A more detailed look was given to the diabetic patients who had a LOS greater than 8 days, as they are the main target of this study.

Tab.1: *Univariate analysis considering patients with LOS greater than 8 days, from the 2019 and 2021 SPARCS NY datasets*

Variables	2019	2021
Age	—	—
0-17	33 (0.4%)	30 (0.4%)
18-29	121 (1.4%)	124 (1.6%)
30-49	1072 (12.6%)	952 (12.2%)
50-69	4252 (49.9%)	3892 (49.8%)
70+	3023 (35.5%)	2810 (35.9%)
Sex	—	—
Male	5297 (62.3%)	2990 (61.7%)
Female	3204 (37.7%)	4818 (38.3%)
Race	—	—
Black	2482 (29.2%)	2361 (30.2%)
Multi-racial	95 (1.1%)	68 (0.8%)
White	3800 (44.7%)	3399 (43.5%)
Other	2124 (24.9%)	1980 (25.4%)
Severity	—	—
Extreme	2110 (24.8%)	1654 (21.2%)
Major	4978 (58.5%)	4003 (51.3%)
Minor	142 (1.7%)	173 (2.2%)
Moderate	1271 (14.9%)	1978 (25.4%)
Complications	—	—
Yes	8497 (99.9%)	7806 (99.98%)
No	48 (0.1%)	2 (0.02%)
Total	**8505 (100%)**	**7808 (100%)**

Tab.2: *Univariate analysis considering NY Hospital Service Areas for patients with LOS>8, from 2019 and 2021 SPARCS NY datasets*

Variables	2019	2021
Hospital Area	—	—
Capital/Adirond	501 (5.9%)	382 (4.9%)
Central NY	475 (5.6%)	452 (5.8%)
Finger Lakes	559 (6.5%)	546 (7%)
Hudson Valley	830 (9.8%)	739 (9.5%)
Long Island	1175 (13.8%)	1105 (14.1%)
NY City	4374 (51.4%)	4059 (51.2%)
Southern Tier	122 (1.4%)	108 (1.4%)
Western NY	464 (5.4%)	412 (5.3%)
Total	**8505 (100%)**	**7808 (100%)**

Multivariate Logistic Regression Models and Variable Selection: Moving to the multivariate risk analysis of prolonged hospitalisation, two GLM and GEE models were computed considering the 2019 NY SPARCS dataset, to investigate the possible predictors of the in-hospital excessive length of stay. The models were intended to evaluate the following covariates:

<u>FULL Model</u>: LOS8 ~ β0 + β1 white + β2 males + β3 d_risky + β4 d_severe + β5 compl + β6 d_admtype + β7 PCI + β8 dialysis + β9 ampt + β10 age4 + β11 age5

After proceeding with the Backward Elimination technique, since in this present model only one covariate was not significantly related to the excessive length of stay, we decided to stop after eliminating the *compl* variable, a dichotomous variable to assess if a diabetic patient had diabetes-related complications or not. The specifications of the reduced model are as follows:

<u>REDUCED Model</u>: LOS8 ~ β0 + β1 white + β2 males + β3 d_risky + β4 d_severe + β5 d_admtype + β6 PCI + β7 dialysis + β8 ampt + β9 age4 + β10 age5

Tab.3: *Predictors of in-hospital excessive length of stay on multivariate analysis (GLM and GEE models)*

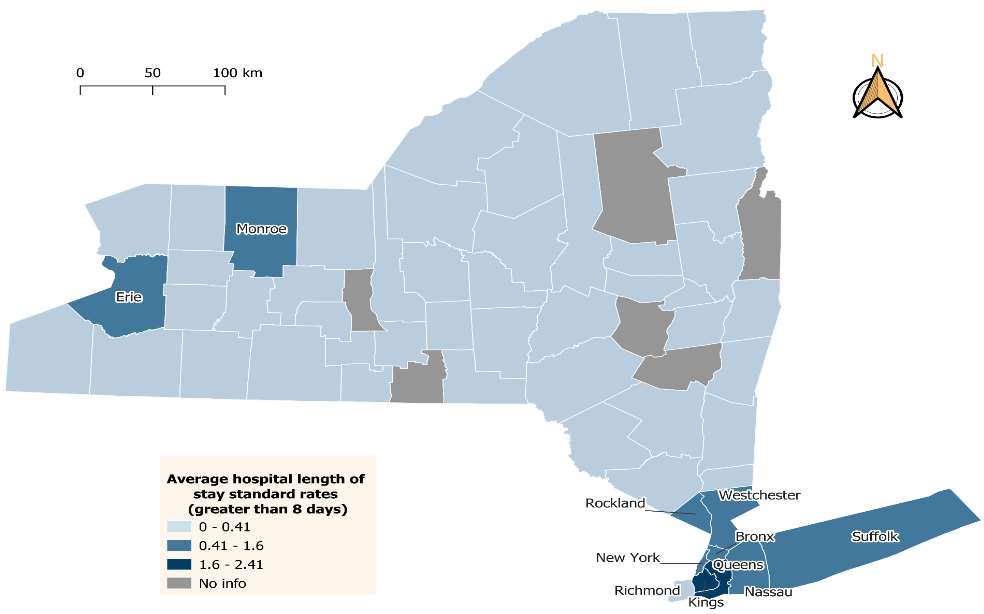
Predictors	GLM Model (OR 95% CI)	P value	GEE Model (OR 95% CI)	P value
white	0.88 (0.83,0.93)	0.000	0.89 (0.84,0.93)	0.000
d_risky	2.04 (1.91,2.17)	0.000	2.03 (1.88,2.11)	0.000
d_severe	3.56 (3.32,3.81)	0.000	3.55 (3.33,3.81)	0.000
males	1.08 (1.02,1.14)	0.004	1.08 (1.03,1.14)	0.005
ampt	4.08 (3.81,4.36)	0.000	4.05 (3.78,4.34)	0.000
d_admtype	1.29 (1.15,1.44)	0.000	1.30 (1.15,1.46)	0.000
PCI	2.25 (1.20,4.13)	0.001	2.23 (1.14,4.36)	0.02
dialysis	0.55 (0.47,0.64)	0.000	0.55 (0.47,0.64)	0.000
age4 (50-69)	1.96 (1.82,2.11)	0.000	1.95 (1.81,2.06)	0.000
age5 (70+)	1.86 (1.72,2.02)	0.000	1.85 (1.71,2.01)	0.000

Table 3 compares the results obtained by applying the reduced logistic GLM and GEE models. By analysing the odds ratios (ORs) and considering the effect of each covariate independently from the others, it is possible to observe that for the upper age classes, LOS8 risk is higher. In comparison with the reference category (0-49), being in the 50-69 age class is 1.96 times riskier, and 1.86 times for 70+. For this reason, it is possible to say that age class is a risk factor for excessive length of stay. We also observed that for both models major or extreme severity, major or extreme risk, being male, undergoing major amputation, being admitted to the hospital as emergency or urgency and undergoing PCI represent statistically relevant risk factors that can increase the length of stay in diabetic patients. On the other hand, being white and undergoing dialysis seem to represent statistically significant protective factors against prolonged in-hospital length of stay for diabetic patients. Since we are comparing two nested models, our approach to model selection involved evaluating their respective log-likelihood function estimates: lnL_F=-18011.8 and lnL_R= -18012. By conducting a log-likelihood ratio test, we obtained LRT=0.25, with a non-significant p-value of 0.625. Despite the proximity of the log-likelihood functions, we opted for the second model due to its fewer parameters.

Furthermore, ROC analysis revealed that both the full and reduced models achieved an accuracy of 82.3%, which is quite good. Thus, in terms of accuracy, both models offer the same probability of correct classification. Both models are also characterised by the same area under ROC curves (AUC), 0.779, a common metric for evaluating model performance. The GEE model that we estimated is characterised by an exchangeable correlation structure, meaning that the within-cluster correlation (alpha) is the same for all clusters. Our group obtained an estimate of alpha equal to 0.047 with a standard error of 0.006.

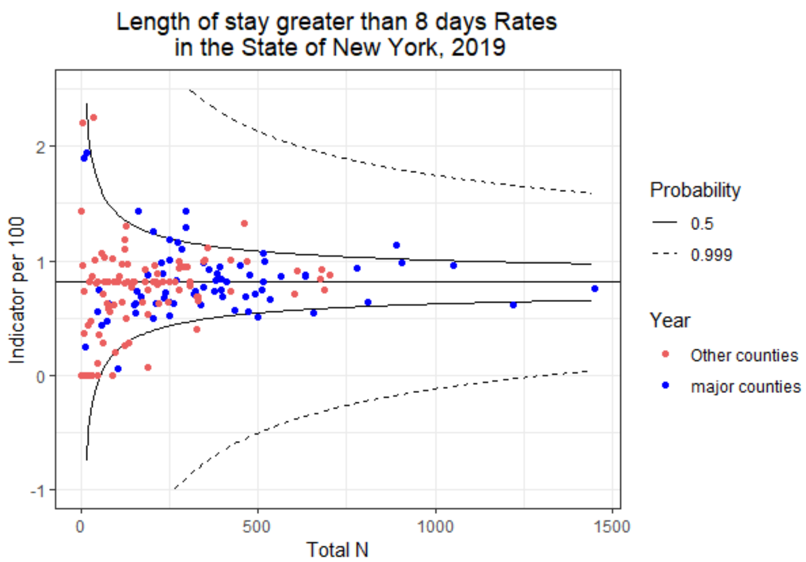
Standardisation Methods: Our analysis also aimed to explore the geographical variation of high lengths of stay across different counties in New York State. Through mapping and funnel plot analysis, we gained insights into the variation of LOS8 rates and identified potential patterns and outliers between the hospital data.

Fig.1: *New York State Counties-Average standard rates for hospital length of stay greater than 8 days, 2019*



To address the geographical analysis, we created a map using Geographical Information Systems and official state shapefiles. The map visualisation revealed distinct geographical patterns in LOS8 rates across New York counties. Specifically, counties in the southern region, notably Long Island, exhibited higher LOS8 rates, correlating with denser population centers. Counties with the highest LOS8 rates (greater than 1.6) included Suffolk, Queens, Kings, Bronx, Manhattan, Richmond, Nassau, Rockland, Westchester, Monroe, and Erie. This spatial analysis suggests a potential association between population density and prolonged length of stay.

Fig.2: *New York State - Funnel plot standard rates for hospital length of stay greater than 8 days, 2019*



The funnel plot analysis highlighted hospitals with LOS8 rates deviating significantly from the average. Notably, hospitals located in major population centers such as Kings, Queens, Manhattan, Suffolk, Bronx, and Nassau exhibited outlier status, indicating potential disparities in healthcare provision or patient management strategies. Further examination of these outliers may elucidate underlying factors contributing to prolonged hospital stays and inform targeted interventions to improve patient outcomes.

To assess COVID-19 differences: To assess differences in prolonged length of stay associated with COVID-19, we examined data from both the 2019 and 2021 SPARCS NY datasets. We created a dichotomous variable that takes into account whether a diabetic patient was discharged in 2019 or 2021. We added this covariate to the existing GLM and GEE models and, in both cases, it resulted statistically significant. Its coefficients, equal to 0.1882 in the former model and 0.1773 in the latter one, lead to a small but significant increase in the likelihood that these patients will have a length of stay longer than 8 days in 2021 compared to 2019. We observe this difference also when examining the standardised rates for hospital length of stay greater than 8 days. For 2019 we have a mean value of 0.793 compared to 0.895 for 2021.

Furthermore, the median values for the two years are 0.43 and 0.59 respectively. This increase can be associated with the COVID-19 pandemic for various reasons. It has strained healthcare systems, both with an increase in patient volumes and with staffing shortages. There could have been a prioritisation of COVID-19 patients in hospitals. Some diabetic patients may have delayed seeking medical care during the pandemic due to concerns about exposure to COVID-19, experienced lifestyle changes, or had reduced access to routine medical care. All of these complications could have resulted in worsening the conditions of the patient, thus resulting in a prolonged length of stay at the hospital.

Conclusions

The exploratory analysis of the dataset provided valuable insights into the demographic and hospital characteristics of diabetic patients with prolonged hospital stays. Through univariate and multivariate analyses, significant predictors of excessive length of stay were identified, including age, severity of condition, gender, amputation status, admission type, and medical procedures undergone. Our findings underscore the importance of considering these factors when managing diabetic patients to optimise hospital resource utilisation and patient outcomes. Moreover, the comparison of full and reduced logistic regression models revealed comparable accuracy and performance metrics, with the reduced model being favoured for its simplicity. These insights enhance our understanding of the factors influencing prolonged hospital stays among diabetic patients and provide a foundation for targeted interventions aimed at improving healthcare delivery and patient management strategies.

Furthermore, our geographical analysis highlighted distinct patterns in prolonged hospital stays across New York counties, with higher rates observed in densely populated areas. Funnel plot analysis identified hospitals with outlier status, suggesting potential disparities in healthcare provision or patient management practices. Investigating these outliers further could elucidate underlying factors contributing to prolonged hospital stays and guide interventions to address healthcare disparities and improve patient outcomes. Additionally, our examination of COVID-19-related differences revealed a small but significant increase in prolonged length of stay among diabetic patients discharged in 2021 compared to 2019. This finding underscores the impact of the COVID-19 pandemic on healthcare delivery and patient outcomes, highlighting the need for adaptive strategies to mitigate the effects of future crises on diabetic patient care. Future studies should also explore how pandemic-related disruptions affect diabetic patient care comprehensively, including the role of telemedicine, access to routine medical care, and strategies to mitigate delays in treatment among vulnerable populations.

References

- Emily Miao, Kaleena Zhang, Jianyou Liu, Juan Lin, Donna Yoo, Claudene J George (2022) Metformin use and mortality and length of stay among hospitalized patients with type 2 diabetes and COVID-19: A multiracial, multiethnic, urban observational study
- Bilal Katipoglu, Mehmet Ilkin Naharci, Eray Serdar Yurdakul (2022) Risk factors predicting hospital length of stay in older patients with type 2 diabetes with Covid-19
- Bala, C., Rusu, A., Ciobanu, D., & Roman, G. (2022). Length of Hospital Stay, Hospitalization Costs, and Their Drivers in Adults with Diabetes in the Romanian Public Hospital System. *International Journal of Environmental Research and Public Health*, *19*(16), 10035.
- Penescu, M., & Mandache, E. (2010). The value of kidney biopsy in diabetes mellitus. *Romanian journal of morphology and embryology= Revue roumaine de morphologie et embryologie*, *51*(1), 13-19.
- Singh, M., Arora, R., Kodumuri, V., Khosla, S., & Jawad, E. (2011). Coronary revascularization in diabetic patients: current state of evidence. *Experimental & Clinical Cardiology*, *16*(1), 16.
- Blonde, L., Umpierrez, G. E., Reddy, S. S., McGill, J. B., Berga, S. L., Bush, M., ... & Weber, S. L. (2022). American Association of Clinical Endocrinology clinical practice guideline: developing a diabetes mellitus comprehensive care plan—2022 update. *Endocrine Practice*, *28*(10), 923-1049.
- Cheng, S. W., Wang, C. Y., & Ko, Y. (2019). Costs and length of stay of hospitalizations due to diabetes-related complications. Journal of diabetes research, 2019.
- Alicic, R. Z., Rooney, M. T., & Tuttle, K. R. (2017). Diabetic kidney disease: challenges, progress, and possibilities. Clinical journal of the American Society of Nephrology, 12(12), 2032-2045.