Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 251 (2024) 696–702

Procedia Computer Science

www.elsevier.com/locate/procedia

The 5th International Workshop on Healthcare Open Data, Intelligence and Interoperability
October 28-30, 2024, Leuven, Belgium

Primary Health Care Appointments and Hospital Stay: An Impact Analysis

Joana Lopes[a], Tiago Miranda[a], Regina Sousa[a,*], José Machado[a]

[a]*LASI/ALGORITMI, University of Minho, Braga 4715, Portugal*

Abstract

The study of avoidable hospitalizations has gained international prom-inence due to its potential to assess the performance of health-care systems. In Canada and Spain, these hospitalizations are analyzed through Ambulatory Care Sensitive Conditions (ACSC), indicating situations that could have been pre-vented or treated without hospitalization. In Portugal, this concept is represented by the term ICSCSP, focusing on care provided in Primary Health Care (PHC). The data analysis in this study aims to determine the impact that medical ap-pointments at PHC may have on the number of hospitalizations, namely, to de-termine whether where the number of medical appointments is greater, the num-ber of hospital stays is lower. During the COVID-19 pandemic in Portugal, many hospitalizations of the elderly were due to the decompensation of chronic diseases, highlighting the importance of access to PHC during health emergen-cies. Data pre-processing was carried out using the Pandas library in Python, merging two datasets monitoring the evolution of hospitalizations and medical appointments in PHC. Despite some challenges encountered during the analysis, such as population bias in district comparisons and the need to adjust metrics to properly reflect the relationship between appointments and hospital stays, it was concluded that the number of appointments in PHC does not have a direct im-pact on hospitalizations. For a more accurate analysis, it would be necessary to consider other factors, such as patient and district characteristics, and conduct more targeted studies, especially after disruptive events like the COVID-19 pan-demic. This more detailed analysis would allow for a better understanding of the relationship between medical appointments and hospitalizations, contributing to improvements in the healthcare system.

Keywords: Primary Healthcare, Aboidable Hospitalizations, Ambulatory Care

* Corresponding author. Tel.: +351 253604430
E-mail address: regina.sousa@algoritmi.uminho.pt

1877-0509 © 2024 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the Conference Program Chairs
10.1016/j.procs.2024.11.171

1. Introduction

Avoidable hospitalizations have gained significant attention as they are associated with the performance of healthcare systems. In other words, since avoidable hospitalizations represent situations that could have been prevented (primary prevention and health promotion at the primary healthcare level) or treated without the need for hospitalization (secondary or tertiary prevention at the outpatient care level), the occurrence of hospitalization may indicate a failure in the healthcare system regarding prevention [1]. This topic becomes particularly relevant when the number of avoidable hospitalizations translates into significant healthcare expenditures and an increase in the mortality rate. The objective of this study is to determine whether, in regions where the number of medical appointments provided to patients in the context of Primary Health Care (PHC) is higher, there is a lower number of hospital stays. Through this analysis, we aim to contribute to the identification of avoidable hospitalizations as a topic of great importance within the National Health System (NHS). In this regard, the research question posed is as follows: ”Does the number of medical appointments provided under PHC have an impact on the number of hospital stays?”

2. State of the Art

The study of avoidable hospitalizations has gained special relevance internationally, particularly in Canada and Spain, where it’s possible to quantify their financial impact on healthcare. In these countries, through the analysis of ACSC, a set of pathologies for which there is the possibility of prevention or treatment without resorting to the hospital, it’s possible to assess the performance of healthcare systems. Translating the acronym into Portuguese generates a more comprehensive concept since it involves outpatient care that may not be solely linked to PHC. The few studies on this topic use the acronym ICSCSP, “Internamentos por Causas Sensíveis aos Cuidados de Saúde Primários” (Hospitalizations due to Conditions Sensitive to Primary Health Care), considering that this way it encompasses only the care provided in PHC, namely at the level of primary prevention, early diagnosis and treatment, and chronic disease management [1]. Although the pathological cause that motivates avoidable hospitalizations is a crucial factor for this analysis, not only does this factor influence these hospitalizations. Demographic and socioeconomic characteristics should also be considered when trying to understand what other conditions, not just pathology, characterize the people who are hospitalized. Among pathological causes, chronic diseases emerge as the most frequent reason for hospitalization in elderly individuals [2]. Given that in Portugal, there is no clear definition of ICSCSP, some studies have tried to adapt the methodology used in other countries to conclude the impact of avoidable hospitalizations on the Portuguese NHS. A study conducted in 2020 revealed that according to the Spanish methodology, the total cost of avoidable hospitalizations was €149,747,010.64 [3]. Despite the National Health Plan, a document that diagnoses and outlines interventions for health in Portugal, highlighting concerns about hospitalization rates for chronic diseases, and preventable diseases by vaccination, among others, we consider, given the scarcity of validated studies for the Portuguese reality, that this issue lacks awareness [4, 5]. Alongside this, if we consider the concept of population aging currently experienced in Portugal, we realize the dimension this issue can take. As nurses, according to our lived experience in the post COVID pandemic era, we note that many of the reasons for hospitalization in the elderly age group were due to decompensations of chronic diseases, namely Chronic obstructive pulmonary disease (COPD), Congestive heart failure (CHF) and diabetes. Knowing that, at the peak of the pandemic, access to PHC was very limited, we can say that the lack of access to PHC may, in some cases, have led to hospitalization. In others, given the increased adherence of people to the SNS24 hotline, which was previously ”dormant”, this resource may have contributed to avoiding some emergency department visits and possible subsequent hospitalizations. There are still no major studies on the full impact of COVID-19; however, we consider that concerning avoidable hospitalizations, it could be interesting to analyze data related to the reason for hospitalization of patients, knowing that access to primary and secondary prevention was limited.

3. Materials

Both datasets were extracted directly from reliable sources, namely certified web-sites such as the Portuguese Public Administration’s open data portal (dados.gov) and the “Transparency” area of the Portuguese National Health Service (“Sistema Nacional de Saude” - SNS) website, an Open Data initiative carried out by the Ministry of Health, in order to make available the vast set of data that underlie the operations and transactions that take place within the scope of the SNS's activities. This way we guarantee the reliability of the data.

3.1. Dataset 1: "atividade-de-internamento-hospitalar"

The dataset " atividade-de-internamento-hospitalar" [6], monitors the evolution of the number of discharged patients and the length of hospital stays per hospital institution and date (year and month), from January 2015 to December 2023. This dataset comprises 14,215 records and 7 distinct columns that organize the data by period, region, institution, geographical location, specialty type, discharged patients, and days of hospitalization. It is worth noting that the column "Doentes Saídos" is an indicator that considers all patients discharged from a given institution (excluding internal transfers), meaning we can assume that the number of discharged patients will approximately correspond to the number of hospital stays and this column will be renamed "Número de internamentos". A more detailed description can be found in the supplementary document to this article

3.2. Dataset 2: "evolucao-das-consultas-medicas-nos-csp"

The dataset " evolucao-das-consultas-medicas-nos-csp" [7], monitors the monthly evolution of in person, remote/unscheduled, and home medical appointments in PHC, by ACES "Agrupamento de Centros de Saúde" (Health Center Grouping), from January 2014 to December 2023. This dataset comprises 6,900 records and 7 distinct columns that organize the data by period, region, entity, geographical location, number of in-person medical appointments, number of remote or unscheduled medical appointments, and number of home medical appointments. A more detailed description can be found in the supplementary document to this article.

4. Methods

This work followed ETL (Extract-Transform-Load) [8] principles, and the data extraction phase has already been explored in the previous section. At this stage we transform our data for subsequent loading and storage in the database. To manipulate our data, we used the Pandas library in Python, which allowed us to efficiently transform the data and prepare it for analysis. This included removing irrelevant columns, eliminating missing values, and renaming columns and rows with eligible characters. The most important step of this preprocessing, which enabled us to merge the two datasets, was creating a mapping for each dataset. The mapping allowed us to assign the institutions from dataset 1 [6] and the entities from dataset 2 [7] to the corresponding district, and in the end, add a "Distrito" column to each dataset. After that, we needed to aggregate our raw data: in dataset 1, we aggregated the duration number of hospital stays and the number of hospital stays by "Região" "Período" and "Distrito", and in dataset 2, we aggregated the number of presential medical appointments, re-mote/unscheduled appointments, and home appointments by "Região" "Período" and "Distrito". This way, we obtained a single entry for each period/region/district in each dataset, and then we merged the two datasets using these same columns. We found that Pandas performed more efficiently than PySpark for merging the datasets. This might be because the final dataset is not very large and can easily fit into the available RAM. In such cases, Pandas can be faster as it avoids the overhead of distributing data across a cluster [9]. Regarding data loading, the dataset with processed and ready-to-analyze data was loaded into MongoDB [10] using a Python script with PyMongo [11]. Finally, the tool used for data analysis and visualization was Power BI, which allowed us to create valuable interactive dashboards for the data interpretation stage. In Power BI, when we began analyzing the data, we encountered some difficulties. Specifically, the comparison between districts regarding the number of medical appointments and the number of hospital stays was biased by the population differences between districts. Therefore, we needed to import a dataset, "População Residente" [12], which allowed us to adjust the data according to the number of inhabitants in each district. Additionally, we noticed that the "Número de internamentos" column represented cumulative monthly values of hospital stays, meaning the value for February included the value of that month plus January. Thus, to calculate the metric of hospital stays per inhabitant, we had to add a calculated column "Monthly Hospital Stays," where the value for each month would be the value presented for that month minus the value of the previous month. Other metrics were generated to build the dashboards that enabled us to draw conclusions on this topic, and they can be accessed in the supplementary document to this article. Since the analysis was primarily focused on years rather than specific months with-in each year, we generated a calculated column ”Year,” while still allowing users to filter the data by month.

5. Results

Returning to the initial question that motivated this work, the objective was to understand whether the total number of appointments provided in PHC had a direct im-pact on the number of hospital stays. Specifically, we wanted to determine if the districts with more appointments have fewer hospital stays. Various visuals were generated, allowing us to draw some conclusions on this topic. Analyzing the TOP and BOTTOM visuals, we could immediately see the 5 districts with the highest and lowest numbers of appointments per inhabitant and hospital stays per inhabitant. When analyzing the 5 districts with the highest total number of appointments in PHC and the 5 districts with the lowest total number of hospital stays, we observed that they were not the same (fig.1). For example, the district with the highest number of appointments in PHC is Porto. However, when looking at the districts with the lowest total number of hospital stays, Porto is not in the TOP 5.

Fig. 1. TOP districts with higher Appointment per inhabitant and TOP districts with higher Hospital Stays per inhabitant.

When we examine the 5 districts with the lowest total appointments per inhabitant, we see that some of these districts, namely Leiria, Aveiro, Setúbal, and Viseu, are also the districts with the lowest total hospital stays(fig.2). This indicates that the number of appointments provided in PHC alone does not have a direct impact on the number of hospital stays.

Additionally, the scatter plot presented allows us to more visually verify the conclu-sions mentioned above. The y-axis represents the total number of hospital stays per inhabitant in descending order, meaning the higher up on the y-axis, the lower the number of hospital stays. The x-axis represents the total number of appointments per inhabitant in ascending order, meaning the farther from zero, the more appointments have been given. We can observe that the districts farther to the right on the scatter plot are not necessarily the districts higher up on the y-axis. Looking again at the dis-trict of Porto, we see that although it has the highest ratio of appointments per inhabit-ant (more appointments per inhabitant), many districts above it show a better ratio of hospital stays per inhabitant (fewer hospital stays per inhabitant) (fig.3). The evolutionary graph allows us to see the variation over the years in the total number of hospital stays per inhabitant and the total number of appointments per in-habitant (fig.4). A key point to highlight in the analysis of this visual is the year 2020, where we see a significant decrease in hospital stays alongside an increase in appointments. An unconsidered analysis might lead us to conclude that this increase in appointments explains the decrease in hospital stays. However, we know that in April 2020, the COVID-19 pandemic began in Portugal,

Fig. 2. BOTTOM districts with lower Appointment per inhabitant and BOTTOM districts with lower Hospital Stays per inhabitant

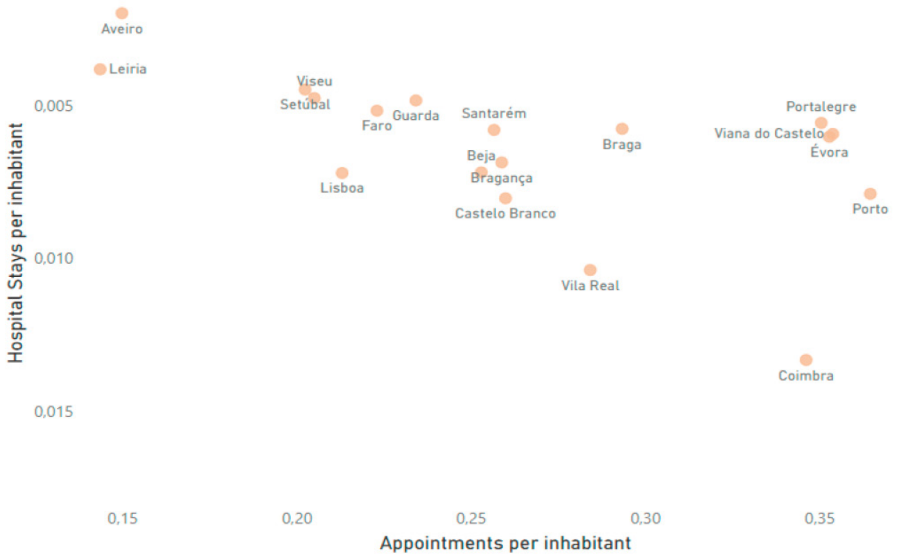

Fig. 3. Scatter plot showing the Hospital stays per inhabitant (y-axis) and the Appointments per inhabitant (x-axis).

leading to population confinement and consequently a reduction in emergency visits and hospital stays. Another interesting aspect when looking at the increase in appointments is the type of appointments given. In 2020 and 2021, years that show increases in appointments in PHC, we can see that most of the appointments were non-presential (remote/unscheduled or home visits), with values around 60% for this type of consultation, reminding us of the incentive to contact the SNS24 helpline for health advice. Only in 2022 did this situation reverse, with appointments becoming predominantly in-person (51%).

Lastly, analyzing the visual that uses the "Appointments per Hospital Stays" ratio along with the interactive geographical analysis visual, we can immediately identify which districts have a higher ratio, meaning districts with a more intense color are those where the number of appointments is much higher than the number of hospital stays. Looking at the district of Aveiro, we see it has the best appointments per hospital stays ratio. However, compared to

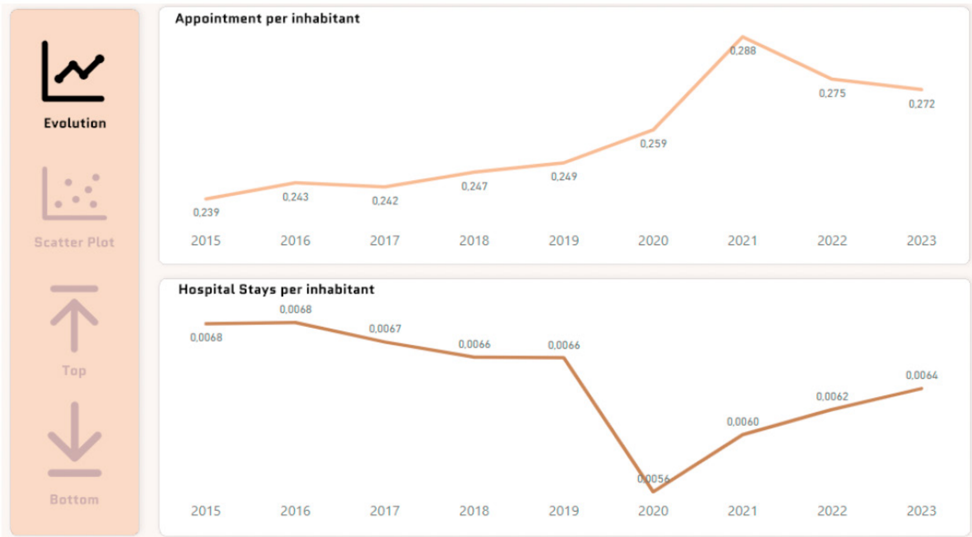

Fig. 4. Evolutionary graph representing the variation over the years in the total number of hospital stays per inhabitant and the total number of appointments per inhabitant.

other districts, it is in the TOP 5 districts with the fewest appointments per inhabitant and also in the TOP 5 districts with the fewest hospital stays per inhabitant, indicating no direct relationship between the variables. Once again, we can reinforce that it is not only the number of appointments given that contributes to the low number of hospital stays, highlighting the need for other factors to justify this complex phenomenon.

6. Discussion and Conclusions

From the analysis of these data, the conclusion we can draw, as an answer to our initial question, is that the number of appointments provided in PHC alone does not have a direct, positive or negative, impact on the number of hospital stays. Although studies in this area are scarce, some publications have highlighted the existence of multiple factors that influence the numbers related to hospital stays. Demographic and socioeconomic characteristics should also be considered when trying to understand the other conditions, not just the pathology, that characterize people who are hospitalized [2]. It is important to remember that the initial datasets were independent and were not originally produced to address the objective outlined at the beginning of this article. Data collection is an essential step for the success of an analysis of this type. At this stage, we realize that many other factors should have been considered in this analysis, namely the characteristics of the hospitalized individuals, the reasons for hospitalization, district-related characteristics, etc. A study more suited to this question would involve targeted data collection, where, for example, districts with similar sociodemographic characteristics are chosen, patients belonging to the 50-80 age group are selected, and a filter on hospitalizations/ appointments is applied. In this case, hospitalizations related to the decompensation of chronic diseases and corresponding appointments aimed at these risk groups would be considered. Regarding trends based on recent years, we can see that alongside a slight increase in the number of PHC appointments since 2015, the number of hospitalizations has been slightly decreasing since that year. However, the COVID-19 pandemic disrupted these gradual variations. Given the scarcity of studies related to this phenomenon, it would be interesting in the future to conduct a comparative study between the years before and after the COVID-19 pandemic, in an attempt to understand whether there has been any increased awareness among people regarding the importance of follow-up in PHC

Acknowledgements:

This work has been supported by FCT (Fundacão para a Ciência e Tecnologia) within the R&D Units Project Scope: UIDB/00319/2020.

References

[1] Sarmento J., Alves, C., Oliveira, P., Sebastião, R., & Santana, R. (n.d.). Caracterização e Evolução dos Internamentos Evitáveis em Portugal: Impacto de Duas Abordagens Metodo-lógicas Characterization and Evolution of Avoidable Admissions in Portugal: The Impact of Two Methodologic Approaches. https://pdfs.semanticscholar.org/0199/2b55b1d91a4577ae5498339fc319b5e235e9.pdf

[2] Tian, Y., Dixon, A., & Gao, H. (2012). Emergency hospital admissions for ambulatory care-sensitive conditions: identifying the potential for reductions DATA Briefing. https://assets.kingsfund.org.uk/f/256914/x/d4cc8d2ce3/data_briefing_emergency_hospital_admissions_2012.pdf

[3] Moreira, M., Pereira, A., Doutora, R., Fonseca, D., Duarte, P., De, S., & Tavares. (2020). Os Internamentos Evitáveis como Agente Modelador do Financiamento dos Cuidados de Saúde no SNS. https://repositorio.ul.pt/bitstream/10451/45318/1/ulfc126012_tm_Mariana_Pereira.pdf

[4] Saúde Sustentável: de todos para todos SAÚDE DA POPULAÇÃO EM PORTUGAL. (n.d.).https://pns.dgs.pt/files/2023/02/PNS2021-2030_Saude-da-Populacao-em-Portugal.pdf

[5] PNS 2021-2030 – Plano Nacional de Saúde. (n.d.). Pns.dgs.pt. https://pns.dgs.pt/pns-2021-2030/

[6] Governo da República Portuguesa - Ministério da Saúde. (2024). Atividade de Internamento Hospitalar — Transparência. https://shorturl.at/cfiwK

[7] ACSS. (2024, February). Consultas Médicas nos Cuidados de Saúde Primários - da-dos.gov.pt - Portal de dados abertos da Administração Pública. https://dados.gov.pt/pt/datasets/consultas-medicas-nos-cuidados-de-saude-primarios-2/#resources

[8] Seenivasan Mphasis, D., & Seenivasan, D. (2023). ETL (Extract, Transform, Load) Best Practices. Article in International Journal of Computer Trends and Technology, 71, 40–44. https://doi.org/10.14445/22312803/IJCTT-V71I1P106

[9] Chauhan, A. (n.d.). A Review on Various Aspects of MongoDb Databases. Retrieved June 6, 2024, from https://www.stackchief.com/search/mongodb

[10] Somasundar, A. V. S. S., Chilakarao, M., Krishnam Raju, B. R., Kumari Behera, S., Rama-na, C. V., & Sethy, P. K. (2024). MongoDB integration with Python and Node.js, Express.js. 2024 4th International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies, ICAECT 2024. https://doi.org/10.1109/ICAECT60202.2024.10469546

[11] Mozzillo, A., Zecchini, L., Gagliardelli, L., Aslam, A., Bergamaschi, S., & Simonini, G. (n.d.). Evaluation of Dataframe Libraries for Data Preparation on a Single Machine. Retrieved June 6, 2024, from https://github.com/dbmodena/bento

[12] PORDATA. (n.d.). População nos Censos em Portugal de 2021 — Pordata. Retrieved May 31, 2024, from https://www.pordata.pt/censos/resultados/populacao-portugal-1075