Miguel Félix Mata-Rivera
Roberto Zagal-Flores
Daniela Elisabeth Ballari
José Antonio León-Borges (Eds.)

Communications in Computer and Information Science 2298

Geographical Information Systems

5th Latin American Conference, GIS-LATAM 2024
Mexico City, Mexico, September 2–6, 2024
Proceedings

Springer

GIS LATAM

Contents

Spatio-Temporal Analysis of Diabetes-Related Hospitalizations: Geographical Areas of Mexico

Carlos Hernandez-Nava[1(✉)], Miguel-Felix Mata-Rivera[1], and Sergio Flores-Hernandez[2]

[1] Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas del Instituto Politécnico Nacional, 07340 Gustavo A Madero, Mexico City, Mexico
{hernandeznc,mmatar}@ipn.mx

[2] Instituto Nacional de Salud Pública de México, 62100 Cuernavaca, Morelos, Mexico
sergio.flores@insp.mx
https://www.upiita.ipn.mx/

Abstract. The area of geographic information system (GIS) is attracting growing attention because geographic data are now everywhere, in this work with electronic health records. Few studies have explored avoidable diabetes-related hospitalizations in Mexico, and there is no systematic process. This paper provides a spatio-temporal analysis through calculating age adjusted rates by sex and geographical area. This study aims to determine the rates of avoidable hospitalizations as a metric to assess the quality of primary diabetes care prevention, stratifying the country in eight geographical areas. We followed the method used in epidemiology and the approach is mainly based on biostatistics techniques and the GIS process. These results emphasize the importance of working together between GIS experts and epidemiologists. The data presented here reveal that the geographic area with the highest rate of avoidable diabetes-related hospitalizations is the peninsula of Mexico. These findings add substantially to our understanding of how diabetes is cared for and decision makers can generate public health policies and improve health care systems.

Keywords: Spatio-temporal · Data science · Diabetes · Avoidable Hospitalizations

1 Introduction

The area of applied data science is receiving increasing attention because there are more and more data and, in 2024, 328.77 million terabytes are being created every day[1]. In the annual reports "Panorama epidemiológico de las enfermedades no transmisibles en México" of 2020 and 2021 of the Secretaría de Salud de México [1,2], it is concluded that:

[1] https://whatsthebigdata.com/data-generated-every-day/ Date: 07/03/2024.

M. F. Mata-Rivera et al. (Eds.): GIS-LATAM 2024, CCIS 2298, pp. 73–85, 2025.
https://doi.org/10.1007/978-3-031-80017-7_5

> ... the importance of [investigating] these conditions and that in the near future we can have a system... that allows us to group together some of the conditions that make up the metabolic syndrome and even go steps back to the early identification of risks in the population located in Mexican territory.

Although the General Epidemiology Department report, 2021 [2], refers to diseases such as type 2 diabetes mellitus that have a persistently high incidence and are in the early stages, which is alarming. With particular reference to Mexico, in the Ministry of Health, there are currently no procedures in place to integrate the databases, clean the data, and thus have them ready to analyze them, obtain models and support decision-making by health specialists.

A health indicator is a measurement that reflects a given situation. Any health indicator is an estimate (measurement with some degree of imprecision) of a given dimension of health in a specific population [3].

Health indicators serve to describe and monitor the health status of a population. Attributes are characteristics or qualities of health. The dimensions of health are physical, emotional, spiritual, environmental, mental, and social well-being. Indicators are dynamic, under specific temporal and cultural contexts and situations [4]. An attribute that most indicators have is the ability to measure them at different geographical levels (regional, national, and local) and population subgroups (age, sex, ethnicity, among others).

Abedjan et al. [5] refers to the existence of a technical challenge or problem when applying data science techniques in health, which is to handle heterogeneous and multidimensional data, with heterogeneity meaning data from different sources, and multidimensionality being the union of the variables or dimensions provided by each of the sources, for example, data from socioeconomic levels, places of residence, atmospheric conditions, and specific health care habits, and concludes that to design reliable models, it is first necessary to integrate, clean, understand and analyze these data.

Among the most commonly used stratification factors for monitoring health inequalities, are included in the acronym PROGRESS, which in English, Place, Race/ethnicity, Occupation, Gender, Religion, Education, Socioeconomic status, Social capital [6]. Taking into account all conditions sensitive to primary care, the Agency for Healthcare Research and Quality specifies the Prevention Quality Indicators (PQI), in this work are used the following diabetes-related indicators [7]:

PQI 01 Diabetes, short-term complications admission rate,
PQI 03 Diabetes, long-term complications admission rate,
PQI 14 Uncontrolled diabetes admission rate,
PQI 16 Lower extremity amputations among patients with diabetes admission rate, and
PQI 93 Prevention Quality Diabetes Composite

The term crude rate has been used to denominate the rate between cases by study population. It is a general measure that does not take into account the diversity of groups in the population, and in this paper is expressed as the rate per 100,000 inhabitants.

The terms adjusted rate and standardized rate are used here interchangeably, which means the rate obtained by dividing the cases of a specific group by the study population. Differences in age and gender structures between populations are adjusted, allowing direct comparisons to be made. Although age is normally used in this process, other factors (e.g., ethnicity, religion, etc.) can also be used. A single statistic is produced that allows easy comparisons between populations.

Throughout this work, an avoidable hospitalization or a potentially avoidable hospitalization will represent a hospitalization related to ambulatory care sensitive conditions (ACSC), that is, if primary or ambulatory care is timely and effective, it is not necessary to go to the hospital [8]. The avoidable diabetes-related hospitalization rate is the rate of cases that match the ICD code listed in the specification PQI 93 by the study population.

This investigation was carried out to analyze the avoidable rates of hospitalization related to diabetes over time (2010–2022) and space (geographical areas of Mexico). In epidemiological methods, tables and classical graphics are commonly used, even geographical variables are used in studies. Therefore, this work also shows how GIS improves the analysis of epidemiological studies, in this case diabetes-related hospitalizations.

2 Related Work

In this section, related works from the USA, Portugal, and Mexico are presented. They filter diabetes cases that match ICD codes with others used in the literature. To compare results between studies across the world, it is necessary to use the same ICD codes, and for diabetes-related ICD codes listed in the PQI specification by AHRQ is the best option to compare with other studies, regional or national.

Shiraz et al. [9] analyzed the records of 31,934 patients from the Primary Care Clinical Health System in the states of Minnesota, Iowa, and Wisconsin in 2019, in the United States of America. A cross-sectional study is designed for analysis. Their aim is to explore the association between diabetes care in rural and urban areas. And they find that patients who are more deprived and live in a rural area are significantly less likely to receive high-quality diabetes care compared to patients living in less deprived and urban areas, and conclude that public health efforts should focus a little more on rural areas.

Ramalho et al. [10] present the PQI diabetes prevention quality indicators: 1, 3, 14, 16 and the composite indicator 93 in Portugal, where they divide into 5 geographic regions. They also evaluated efficiency at the first level of care using PQI for the years 2016–2017. The regression model they use for the evaluation is Tobit or censored regression model. Among the results they mention that 13 out of 54 groups of health centers manage to become efficient and that in large part people under 45 years of age who are controlled are what reach more efficient diabetes care. Among the conclusions, they say that most primary care health centers are inefficient in diabetes prevention, and by addressing this situation, they could improve the rates of avoidable hospitalizations.

Flores-Hernández et al. [11] present an evaluation of the quality of care of patients with diabetes and present its association with the follow-up of glycemic control. The study was carried out through a cross-sectional analysis of the years 2006 (2,965 patients) and 2012 (4,483 patients); in both cases, the patients are 20 years old or older. The source of the data is the National Health and Nutrition Survey for the respective years.

Multiple regression models (logistic and linear) were created to evaluate changes over the years. The estimated change between was 2.1 more in 2012 than in 2006 (95 CI 1.5-2.7). To evaluate quality, Flores-Hernández et al. used the following criteria: glucose monitoring, doctor visits to treat diabetes, detection of hypertension, detection of dyslipidemia, overweight care, detection of protein in urine, detection of retinopathy, detection of diabetic foot, and whether the patient has pharmaceutical treatment. They concluded that diabetes continues to be a public health emergency in Mexico and that strategies are needed to improve the quality of care of patients with diabetes.

Continuing in the investigation of the challenges in evaluating the quality of diabetes care in Mexico [12], similar to the work of the 2015 publication by Flores-Hernandez et al., estimate quality indicators in primary care and analyze association with glycemic control and use the results of the 2012 and 2018–19 National Nutrition and Health Surveys of people aged 20 years or older, a total of 9,038 people.

To measure quality of care, they take four aspects: diabetes control; surveillance of cardiovascular risk factors; early detection of complications; and follow-up with pharmacological treatment. They use multiple linear regression and a logistic regression model when adding variables of glycemic control and overall quality of care of adults with diabetes. They concluded that population-level studies are needed for better results.

The presented studies focus on analyzing the rates of potentially avoidable hospitalizations related to diabetes, one of them using the same indicators that are used in this study, and the others propose aspects such as diabetes control, early detection of complications, and follow-up of pharmacological treatment. In addition, one of them uses a retrospective observational study with a secondary analysis of databases as is done in this work, and the others make use of the results of surveys and design the cross-sectional study.

3 Data Sources Description

Secondary data sources are those that originally collected data for other purposes. The existing data obtained for this work are considered secondary data.

The source of the data is the open health data website of the Mexican Ministry of Public Health for the years 2010–2022. This work takes the cases of hospital discharges related to diabetes of people 20 years or older, in age groups 20 to 44 years, 45 to 64 years and over 65 years, divided into men and women, and is carried out by geographic areas according to the regionalization defined in the methodology of the National Health and Nutrition Survey 2022 and Planning of the Ensanut Continua[2] 2020-2024.

The open data are in CSV format and are downloaded from the Ministry of Health website [14], specifically the Health Information General Department. In the section hospital discharges there are CSV files from 2010 up to now, and some XLS files with the variable descriptions and catalogs used in the databases. Databases have more than 50 variables and have increased over the years.

The variables used in this work are age, sex, principal condition, and entity, where age is an integer, sex is 1 for men and 2 for women, principal condition for which a person has been hospitalized, contains the ICD code, entity in Mexico this area is called state, and in another level many states create a geographical area.

4 Methodology

The experimental design referred to as secondary data analysis was used. Data were collected from open data sources: DGIS/SS and INEGI from 2010 to 2022. To measure the avoidable rates of hospitalization related to diabetes, the specification of the quality indicator of prevention was used. In fact, similar works are using more and more PQI defined by ARHQ.

The methodology includes: Data collection (Electronic Health Records), Cleaning and Preprocessing (Stratification and Outlier Analysis), Modeling (through biostatistical methods and algorithms), Interpretation (with a biological focus) and Presentation (with statistical evidence). The primary advantages of this methodology are that using data science tools increases the speed of the process in some stages. Data were adjusted or standardized using the direct method for standardization.

Data analysis was performed with the software: STATA v18, R Studio and MS Excel. The missing values are less than 1% and were excluded from the analysis.

This work is a retrospective observational study, it is observational because cases of hospitalizations of patients for some primary condition related to diabetes arose spontaneously, and it is retrospective because the data are obtained from a request after the generation of the data. The years of records are June 2010 to May 2022.

4.1 Methods

The method used follows the foundations of public health methodology, especially epidemiology and biostatistics, but is supported by data science tools, such as database merging, data cleaning tools, and application processing. There are two ways to calculate the crude rate; the first, the naive method or the method

[2] Encuesta Nacional de Salud y Nutrición.

without public health knowledge, is to count the cases of diabetes and divide them by the study population, Eq. 1, which is the population of the geographical area in the period of time of the cases.

$$\text{Crude Rate} = \frac{\text{Diabetes Cases}}{\text{Study Population}} \tag{1}$$

The second is through the specific rates, which are the result of dividing the cases of a specific group by the study population of the same group, Eq. 2, to get the crude rate, just multiply the age group specific rate (Spec Rate) by the age group proportion of the study population in decimal percent (Pop Prop) in Eq. 3.

$$\text{Specific Rate} = \frac{\text{Cases by age}}{\text{Population by age}} \tag{2}$$

$$\text{Crude Rate} = \text{Spec Rate } 20{-}44\text{y} \times \text{Pop Prop } 20{-}44\text{y} \tag{3}$$

In this work to standardize the rates, first, we have to define the population and the diabetes cases stratified by sex and age group; for this step, diabetes cases are the result of the pre-processing and cleaning process. The age-adjusted rate by sex is calculated using the Eq. 4,

$$\begin{aligned} \text{Age Adj Rate} = &(\text{Spec Rate } 20{-}44\text{y} \times \text{Std Pop } 20{-}44\text{y}) \\ &+ (\text{Spec Rate } 45{-}64\text{y} \times \text{Std Pop } 45{-}64\text{y}) \\ &+ (\text{Spec Rate } 65{+} \times \text{Std Pop } 65{+}) \end{aligned} \tag{4}$$

where, Spec Rate 20–44y is the specific rate for population from 20 to 44 years old, Std Pop 20–44y is the proportion of the standard population from 20 to 44 years old; and the same for the groups from 45 to 64 years old and the group from 65 years old and over. That process is a direct standardization [13], and it is applied when we have an age-structured standard population. And is essential to do because the characteristics of the population are different in each geographical areas, because of its eating habits, housing access and quality of health services, and a comparison of crude rates would misleading.

5 Results

The results presented in this section reveal that the southeast of Mexico has higher rates of avoidable hospitalization related to diabetes than other geographical areas of the country. The geographical area of the peninsula of Mexico and the surrounding areas are the most affected by hospitalizations for diabetes and is related to the geographical areas with the highest rate of obesity in Mexico. According to the results of the 2018 national health and nutrition survey, Tabasco and Quintana Roo had the highest percentages of obesity [?].

Table 1 shows the geographical area of Frontera, which is the area with the lowest adjusted rates over the years, the rates are listed per year, and for women and men, also the crude rate, the lower confidence interval (CI low), the upper confidence interval (CI upp) and the error are shown.

Table 1. Frontera: Crude and Age Adjusted Rates by Sex.

Year	Sex	Crude Rate	Adj Rate	CI low	CI upp	Error
2010	Women	33.52	36.89	35.00	38.79	0.97
2010	Men	24.33	27.31	25.62	29.01	0.86
2011	Women	33.45	36.33	34.48	38.17	0.94
2011	Men	25.26	27.98	26.30	29.65	0.86
2012	Women	36.45	38.72	36.86	40.58	0.95
2012	Men	27.90	30.63	28.91	32.35	0.88
2013	Women	34.93	36.66	34.88	38.43	0.91
2013	Men	26.88	29.10	27.46	30.75	0.84
2014	Women	36.09	37.51	35.74	39.28	0.90
2014	Men	27.54	29.55	27.92	31.18	0.83
2015	Women	48.12	49.34	47.35	51.33	1.01
2015	Men	41.79	44.06	42.12	46.01	0.99
2016	Women	33.04	33.53	31.92	35.14	0.82
2016	Men	26.86	28.10	26.57	29.63	0.78
2017	Women	26.70	26.83	25.41	28.25	0.72
2017	Men	21.18	22.01	20.67	23.34	0.68
2018	Women	34.38	34.30	32.72	35.88	0.81
2018	Men	37.06	37.57	35.87	39.27	0.87
2019	Women	30.88	30.54	29.07	32.01	0.75
2019	Men	31.86	32.43	30.86	34.00	0.80
2020	Women	19.72	19.37	18.22	20.53	0.59
2020	Men	20.59	20.75	19.52	21.99	0.63
2021	Women	19.91	19.39	18.26	20.53	0.58
2021	Men	22.32	22.32	21.06	23.58	0.64
2022	Women	25.20	24.38	23.12	25.64	0.64
2022	Men	27.87	27.61	26.23	28.99	0.71

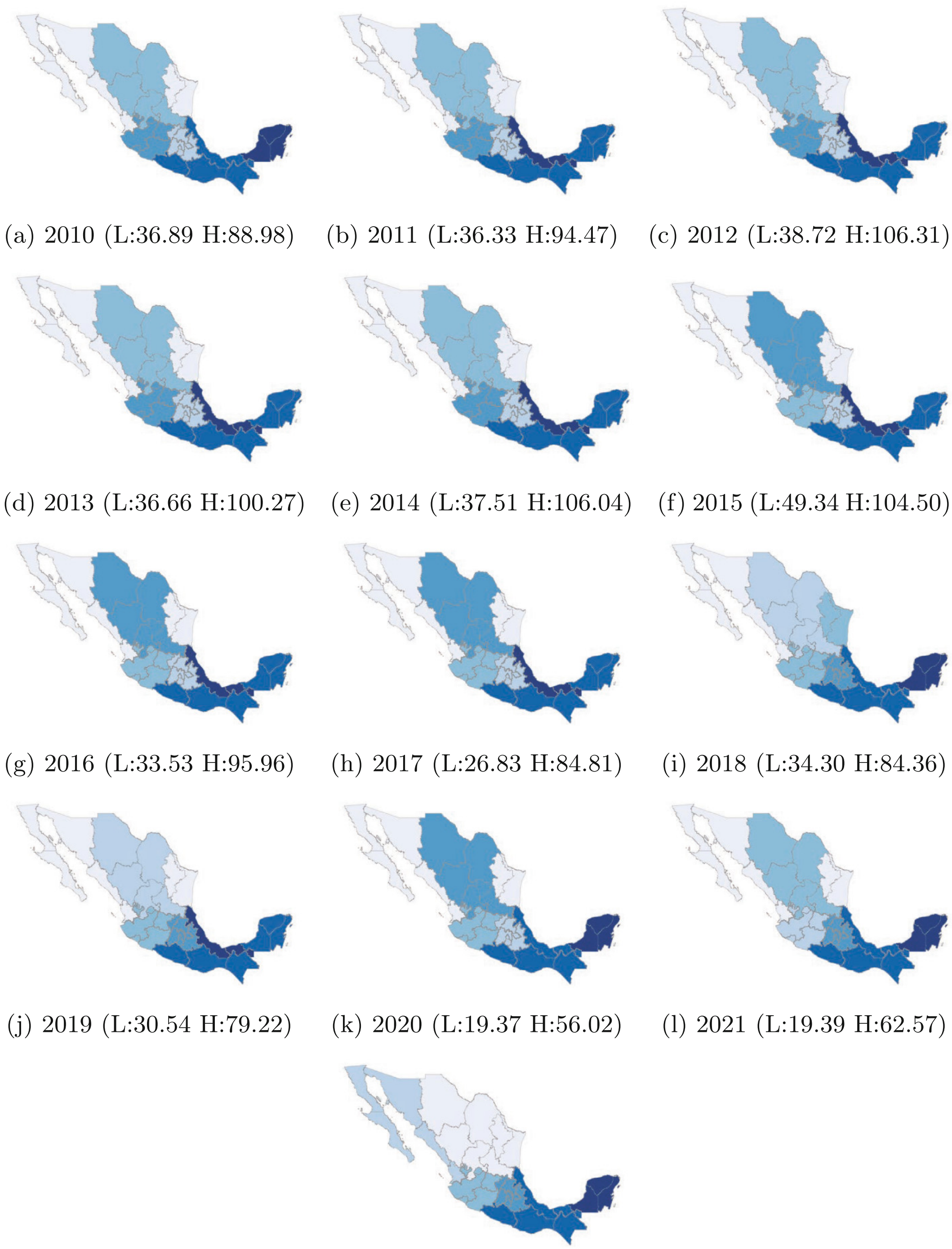
Table 2 shows the geographical area of the peninsula of Mexico, which is the area with the highest adjusted rates, for women and men. The highest rate was in 2012, 144 women were reached in the hospital, although these hospitalizations could have been avoided because diabetes is an ambulatory care sensitive condition (ACSC).

To observe the whole panorama of avoidable diabetes-related hospitalizations in Mexico, Fig. 1 shows the spatial distribution of adjusted rates in Mexico over the years 2010–2022. In parentheses are the lowest and highest rates per year; it can be seen that in the COVID-19 pandemic the highest rates decreased from 79 in 2019 to 56 in 2020 and 62 in 2021, these are values for women.

Table 2. Peninsula: Crude and Age Adjusted Rates by Sex.

Year	Sex	Crude Rate	Adj Rate	CI low	CI upp	Error
2010	Women	103.78	125.08	120.85	129.31	2.16
2010	Men	76.82	88.98	85.38	92.58	1.84
2011	Women	113.01	133.21	128.96	137.47	2.17
2011	Men	83.07	94.47	90.85	98.09	1.85
2012	Women	125.09	144.72	140.39	149.04	2.21
2012	Men	94.87	106.31	102.55	110.07	1.92
2013	Women	122.93	140.19	136.03	144.36	2.13
2013	Men	90.25	100.27	96.69	103.86	1.83
2014	Women	125.13	140.87	136.78	144.97	2.09
2014	Men	96.52	106.04	102.42	109.66	1.85
2015	Women	116.66	129.90	126.04	133.76	1.97
2015	Men	95.77	104.50	100.96	108.03	1.80
2016	Women	109.81	120.86	117.21	124.51	1.86
2016	Men	88.71	95.96	92.63	99.29	1.70
2017	Women	102.96	111.88	108.44	115.33	1.76
2017	Men	79.64	84.81	81.74	87.88	1.57
2018	Women	95.81	103.36	100.09	106.62	1.66
2018	Men	79.50	84.36	81.34	87.38	1.54
2019	Women	87.57	93.65	90.59	96.70	1.56
2019	Men	75.10	79.22	76.33	82.11	1.47
2020	Women	51.60	54.63	52.33	56.92	1.17
2020	Men	53.47	56.02	53.63	58.41	1.22
2021	Women	60.43	63.51	61.07	65.95	1.24
2021	Men	60.02	62.57	60.07	65.06	1.27
2022	Women	72.66	75.88	73.25	78.51	1.34
2022	Men	66.10	68.41	65.84	70.98	1.31

Regarding men, Fig. 2 shows the spatial distribution of diabetes-related hospitalizations, where it can be seen that the southeast of Mexico has the highest rates.



(a) 2010 (L:36.89 H:88.98) (b) 2011 (L:36.33 H:94.47) (c) 2012 (L:38.72 H:106.31)

(d) 2013 (L:36.66 H:100.27) (e) 2014 (L:37.51 H:106.04) (f) 2015 (L:49.34 H:104.50)

(g) 2016 (L:33.53 H:95.96) (h) 2017 (L:26.83 H:84.81) (i) 2018 (L:34.30 H:84.36)

(j) 2019 (L:30.54 H:79.22) (k) 2020 (L:19.37 H:56.02) (l) 2021 (L:19.39 H:62.57)

(m) 2022 (L:24.38 H:68.41)

Fig. 1. Spatio-temporal distribution of Avoidable Diabetes-related Hospitalizations by Geographical Area Of Mexico for Women (L refers to the lowest rate $\times$ 100,000 hab. or clearest area and H refers to the highest rate $\times$ 100,000 hab. or darkest area).

(a) 2010 (L:24.2 H:84.1)

(b) 2011 (L:22.6 H:87.7)

(c) 2012 (L:27.7 H:90.3)

(d) 2013 (L:25.2 H:81.6)

(e) 2014 (L:24.6 H:89)

(f) 2015 (L:24.9 H:87.1)

(g) 2016 (L:22 H:81)

(h) 2017 (L:21.3 H:78)

(i) 2018 (L:31.5 H:81.4)

(j) 2019 (L:30.1 H:68.9)

(k) 2020 (L:20 H:57.1)

(l) 2021 (L:20.4 H:59.8)

(m) 2022 (L:29.2 H:59.4)

Fig. 2. Spatio-temporal distribution of Avoidable Diabetes-related Hospitalizations by Geographical Area Of Mexico for Men.

6 Conclusions

The results of this study reveal that hospitalization rates related to diabetes could be a metric to measure the quality prevention of diabetes, which means that the primary care level could improve diabetes treatments in early symptoms.

Of eight geographical areas, only one is above the OCDE average of avoidable hospitalizations related to diabetes, this is the peninsula of Mexico. In Fig. 3 the adjusted rates are shown, women of the peninsula of Mexico present the highest rates over the years.

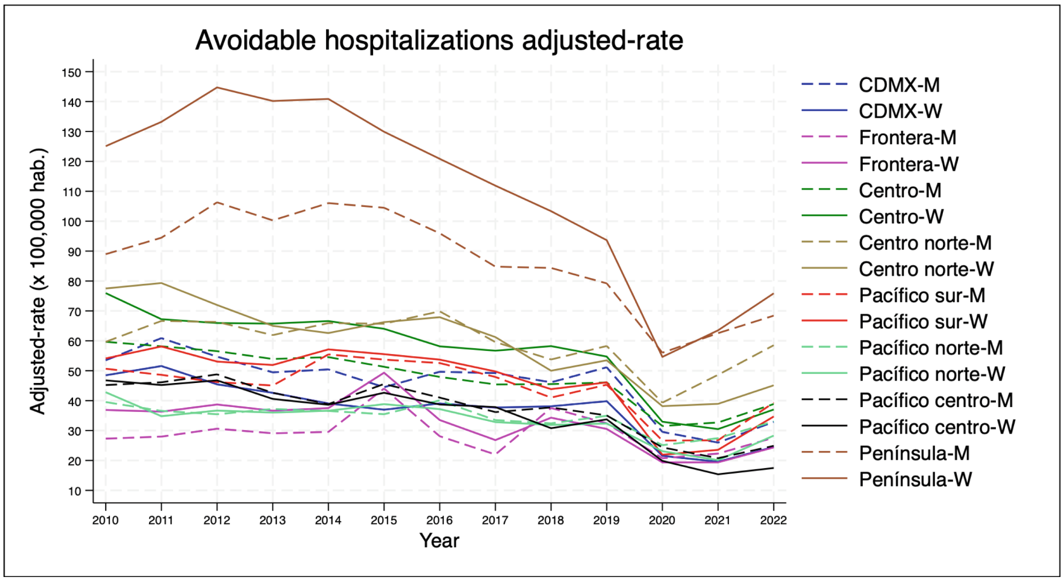


Fig. 3. Avoidable Diabetes-related Hospitalization Rates by Geographical Area.

We have obtained a satisfactory spatio-temporal analysis of avoidable hospitalizations related to diabetes in Mexico stratifying by geographic areas.

The present study has only examined open databases from the public health system of Mexico and does not include the IMSS[3] and ISSSTE[4] databases, two of the main health systems of Mexico, IMSS is the health care system for workers and ISSSTE is the health care system for government workers.

These results provide a significant first step towards a proposal to assess public health care in Mexico at its primary level.

Future research should consider the potential effects of other noncommunicable diseases that are the first cause of death in the world, these diseases could be such as hypertension, cerebrovascular disease, obesity and overweight.

These findings suggest the following recommendations: each geographical area of Mexico must be analyzed per state or small areas, particularly for searching associate factors that could affect or provoke some diseases; another finding

[3] Instituto Mexicano del Seguro Social.

[4] Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado.

is that the mandatory to collaborate interdisciplinary and multidisciplinary, in accordance with the related areas of knowledge; another contribution is to provide data and information for those who make decisions and formulate public health policies. This work laid out the process for analyzing diabetes-related hospitalizations that could replicate other countries.

The results shown in the tables and figures of this work were calculated with STATA v18, which is the software we recommend, and were validated with R Studio scripts and a spreadsheet with formulas written in MS Excel, where the values were in the confidence interval 95%.

Acknowledgments. This study was funded by the CONAHCYT grant.

Disclosure of Interests. The authors have declared that no competing interests exist.

References

1. Panorama epidemiológico de las enfermedades no transmisibles en México 2020, Dirección General de Epidemiología de México. https://epidemiologia.salud.gob.mx/gobmx/salud/documentos/pano-OMENT/panoepid_ENT2020.pdf. Accessed 10 July 2024
2. Panorama epidemiológico de las enfermedades no transmisibles en México 2021, Dirección General de Epidemiología de México. https://www.gob.mx/cms/uploads/attachment/file/745354/PanoEpi_ENT_Cierre2021.pdf. Accessed 10 July 2024
3. Dicker, R.: Principles of Epidemiology in Public Health Practice. Center for Disease Control and Prevention, CDC (2013)
4. Wolfgang, A., Iris, P.: Handbook of Epidemiology. Springer, Cham (2014)
5. Abedjan, Z., Roller, R., Thoms, J: Data science for healthcare: methodologies and applications. In: Consoli S., Recupero R., Petković M. (eds.). Springer (2019). https://doi.org/10.1007/978-3-030-05249-2
6. Emma L., et al.: Using PROGRESS-plus to identify current approaches to the collection and reporting of equity-relevant data: a scoping review. J. Clin. Epidemiol. **163**, 70–78 (2023). https://doi.org/10.1016/j.jclinepi.2023.09.017. ISSN 0895-4356
7. Agency for Healthcare Research and Quality: U.S. Department of Health and Human Services: Quality Indicator User Guide: Prevention Quality Indicators (PQI) Composite Measures, v2023 (2023). https://qualityindicators.ahrq.gov/measures/pqi_resources
8. Caminal, J., Morales, M., Sánchez, E., Cubells, M., Bustins, M.: Hospitalizaciones prevenibles mediante una atención primaria oportuna y efectiva. Journal Atención Primaria **31**(1), 6–17 (2003)
9. Kurani, S., Lampman, M., Funni, S., et al.: Association between area-level socioeconomic deprivation and diabetes care quality in US primary care practices. JAMA Netw. Open **4**(12), e2138438 (2021). https://doi.org/10.1001/jamanetworkopen.2021.38438
10. Ramalho, A., Souza, J., Castro, P., Lobo, M., Santos, P., Freitas, A.: Portuguese primary healthcare and prevention quality indicators for diabetes mellitus - a data envelopment analysis. Int. J. Health Policy Manag. **11**(9), 1725–1734 (2022). https://doi.org/10.34172/ijhpm.2021.76