

Laboratory data statistical analysis from CECCS' patients with or without SARS-CoV-2 (COVID-19)

Análisis estadístico de datos de laboratorio de pacientes del CECCS con o sin SARS-CoV-2 (COVID-19)

YAÑEZ-VARGAS Israel†*, DOÑATE-ÁLVAREZ Andrea, QUINTANILLA-DOMINGEZ Joel and AGUILERA-GONZÁLEZ Gabriel

Universidad Politécnica de Juventino Rosas, Departamento de Ingeniería en Redes y Telecomunicaciones.

ID 1st Author: *Juan Israel Yañez-Vargas* / ORC ID: 0000-0001-5749-8442, CVU CONACYT- ID: 295711

ID 1st Coauthor: *Andrea Abigail Doñate-Álvarez* / ORC ID: 0000-0003-4005-8533

ID 2nd Coauthor: *Joel Quintanilla-Domínguez* / ORC ID 0000-0003-2442-2032, CVU CONACYT- ID: 295711

ID 3rd Coauthor: *José Aguilera-González* / ORC ID: 0000-0002-4160-448X, CVU CONACYT- ID: 329447

DOI: 10.35429/JQSA.2020.21.7.9.17

Received September 05, 2020; Accepted December 01, 2020

Abstract

The increase in the number of COVID-19 cases in Mexico has caused an important effort in the analysis of the virus, due to the above, a statistical study of laboratory data of 109 patients from the State Center for Critical Care of Salamanca (CECCS) with or without COVID-19, whose purpose is to analyze in more detail the disease and the reactions in patients. First, a classification was made in men and women, then positive, negative, healthy and deceased patients. It is important to mention that a data cleaning was carried out, in addition to information filling using the mean of the laboratory values, then the clinical data were normalized from 0 to 1, to obtain the BMI of the patients, box plots and graphs days to days of hospitalization with a sample of patients.

COVID-19, Clinical laboratory data, Statistical analysis

Resumen

El incremento en número en casos COVID-19 en México ha provocado un esfuerzo importante en el análisis del virus, debido a lo anterior, se ha propuesto un estudio estadístico de datos de laboratorio de 109 pacientes del Centro Estatal de Cuidados Críticos de Salamanca (CECCS) con o sin COVID-19, cuyo propósito es analizar más a detalle la enfermedad y las reacciones en los y las pacientes. Primero, se realizó una clasificación en hombres y mujeres, posteriormente pacientes positivos, negativos, pacientes sanos y con defunción. Es importante mencionar que se realizó una limpieza de los datos, además de un llenado de información usando la media de los valores de los laboratorios, posteriormente se normalizaron los datos clínicos de 0 a 1, para obtener las gráficas de la clasificación e IMC de los pacientes, diagramas de caja y gráficas de los días de hospitalización de una muestra de pacientes.

COVID-19, Datos clínicos de laboratorio, Estudio estadístico

Citation: YAÑEZ-VARGAS Israel, DOÑATE-ÁLVAREZ Andrea, QUINTANILLA-DOMINGEZ Joel and AGUILERA-GONZÁLEZ Gabriel. Laboratory data statistical analysis from CECCS' patients with or without SARS-CoV-2 (COVID-19). Journal of Quantitative and Statistical Analysis. 2020. 7-21: 9-17

* Correspondence to Author (e-mail: jyanez_ptc@upjr.edu.mx)

† Researcher contributing as first author.

1. Introduction

In December 2019, a new type of coronavirus originated in the city of Wuhan, China. The scientific family called it Severe Acute Respiratory Syndrome (SARS-CoV-2), which is the cause of COVID-19 that today has become official as a global pandemic (Coronavirus disease (COVID-19), 2020).

Given the great challenge of the scientific community in the knowledge of COVID-19, multiple studies have been carried out for its detection and complete analysis since the behavior of COVID-19 in each country or region will depend on many factors, such as susceptibility to diseases, the most common diseases, age, weight, weather and other factors.

The Federal Health Secretariat of Mexico, through the General Directorate of Epidemiology, reported that, as of December 15, 2020, more than 73 million confirmed cases have been recorded worldwide, in addition to more than 1.6 million deaths from COVID-19. In Mexico, there are currently more than 1.4 million confirmed cases, while there are around 114 thousand deaths (Coronavirus - gob.mx, 2020).

Given the above, it is necessary to know the action of the virus in humans, for which it is intended to carry out a statistical study based on clinical laboratory data obtained from the State Center for Critical Care of Salamanca (CECCS), which is a hospital of third level that was converted into the first hospital dedicated to COVID-19 in the state of Guanajuato and is located in the city of Salamanca (Conversion of the CECCS to a COVID-19 center, 2020).

From all the information obtained and processed, it will be possible to understand the severity of the patient, the condition and know the behavior of the virus in the human body. Carrying out a statistical study or having a data analysis will allow to give greater validity to some affirmations made about the virus, in addition to making better decisions based on mathematical analysis with real data. The branches of engineering and medicine have become very involved in recent years, where statistical studies are continuously carried out in order to learn more about phenomena of global importance, therefore, derived from the global pandemic of COVID-19, the efforts in all areas to learn more about the virus.

In the study by K. Liu et al. (2020, p. 1027) showed the results of processing the epidemiological history data, analyzing the clinical characteristics, treatments and the prognosis of 137 patients admitted with COVID-19, they presented the results on the most recurrent symptoms, the most susceptible ages and chest X-ray images of each patient.

According to a new study where a statistical investigation was carried out on clinical and biochemical data of 12 patients infected with COVID-19, to know the changes in the patient through the use of certain medications, likewise a study based on images was added X-ray and tomography (Y. Liu et al., 2020, p. 372).

(Ng et al., 2020) in their research article, they describe 21 patients confirmed as positive for COVID-19, they underwent X-ray and tomography studies to know their evolution, in addition to showing statistical study of data personal symptoms and some clinical results.

In the study by Yun et al. (2020, 96 p.) Carry out a study with clinical data of nucleic acid and blood from 2510 patients infected with COVID-19, Influenza A and Influenza B, showing statistically comparisons of the changes in each of the laboratory studies.

Zhang et al. (2020, p. 1736) presented in their study an analysis of data from 140 patients with SARS-CoV-2 in which information on demographic data, allergies, clinical manifestations, comorbidities, and laboratory data is grouped for subsequent mathematical analysis.

All previous studies and most of those registered worldwide seek to have a broader knowledge of the behavior of the virus in both men and women, but the existence of documentation for Mexico is much less and if it is established that for patients from the state of Guanajuato and admitted to third-level hospitals is more scarce, makes this research of utmost importance to know the evolution of the virus and patient in a sample of the population infected with COVID-19 of Guanajuato.

2. Material and methods

When the patient is admitted to the tertiary level hospital, a data collection is carried out which includes name, age, weight and comorbidities if they have them, to later be admitted to intensive care.

The data to be analyzed correspond to 109 patients from the CECCS hospital, which is a third-level center located in the city of Salamanca, Guanajuato, they were admitted from March 24 to October 20, 2020.

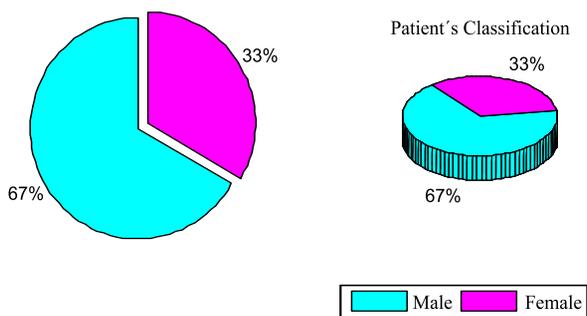


Figure 1 Total patients admitted to the CECCS
Source: Own [Matlab]

Figure 1 shows a diagram of the division of patients admitted to the hospital, of which there are 73 men and 36 women.

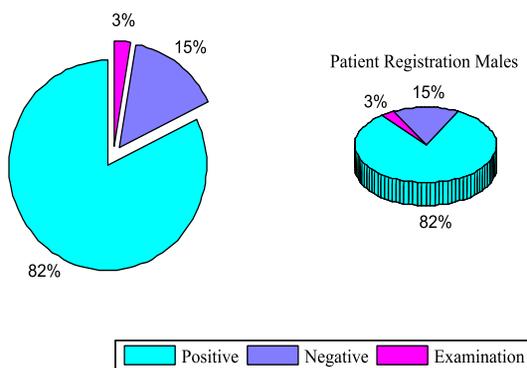


Figure 2 Men admitted to the CECCS
Source: Own [Matlab]

Likewise, in Figure 2 of the 73 men admitted to the hospital, there are 60 patients who tested positive for the Polymerase Chain Reaction (PCR) test for COVID-19, which is the most used due to its great reliability. in the result, while 11 people were negative to the test, leaving 2 people still in observation or waiting for the result of the PCR test (Dra. Celia M. Alpuche Aranda, 2020).

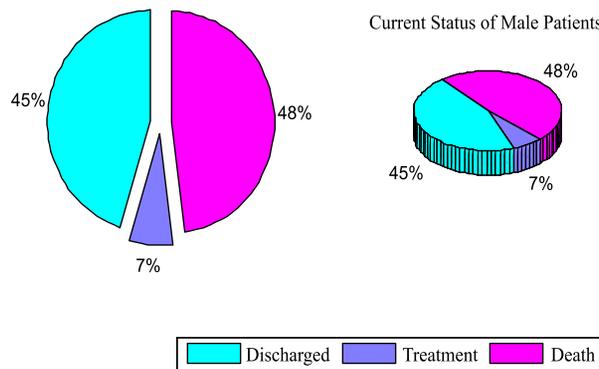


Figure 3 Current status of male patients in the CECCS
Source: Own [Matlab]

From the data shown above, a relationship was obtained of those patients who are still hospitalized with treatment, those who died and those who lived. Figure 3 gives a graphic explanation of the percentage of 109 patients who died, having a high percentage for the number of patients admitted.

In the case of the 36 women who were admitted to the hospital, Figure 4 shows in detail the case of the patients who were positive, negative and who have not yet obtained the result of their PCR test.

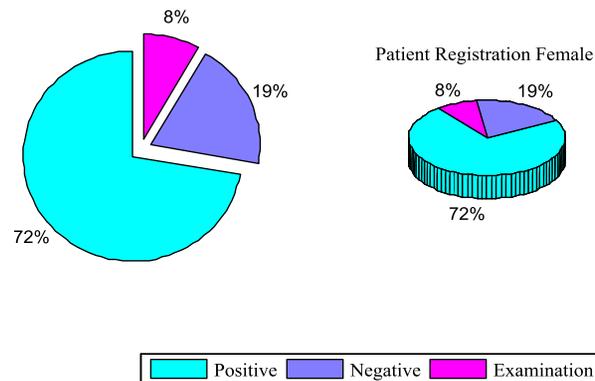


Figure 4 Registry of female patients with their PCR result
Source: Own [Matlab]

Finally, Figure 5 shows results of female patients and their status in the hospital, showing percentages of patients discharged, in death and treatment.

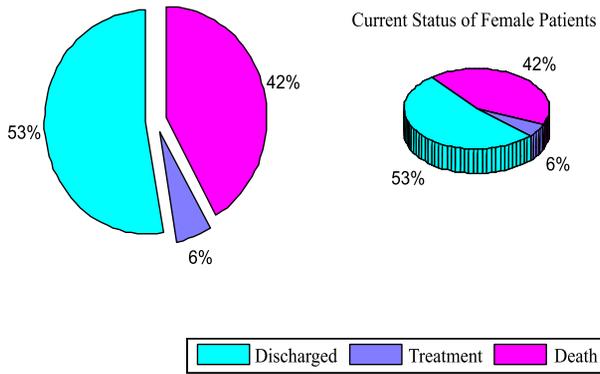


Figure 5 Current status of female patients
Source: Own [Matlab]

As mentioned above, when patients are admitted to the CECCS, measurements of the height and weight of the patients are carried out, to later carry out a study of the Body Mass Index (BMI), either to learn more about the health status of the person and the amount of medication to be prescribed.

Figure 6 describes the BMI of female patients, where it should be remembered that from a BMI value of 25 people become overweight, where it is possible to notice that most women are overweight and some are overweight. grade III obesity.

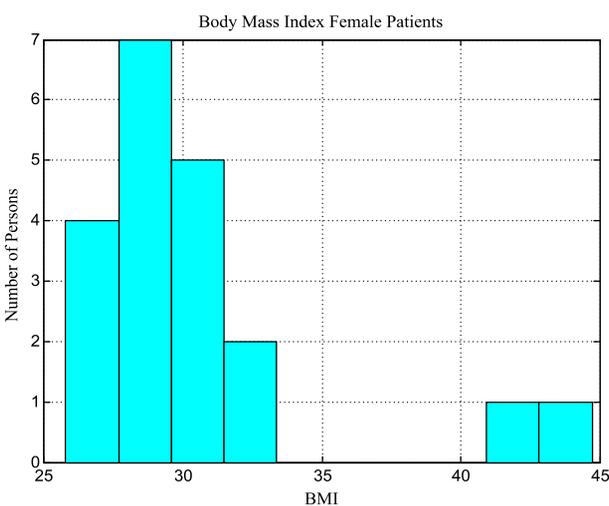


Figure 6 BMI of female patients
Source: Own [Matlab]

In the case of Figure 7, the BMI of men is shown, in which an important difference can be noticed with respect to the graph of women, since men are more overweight than women, despite the fact that the greater cases of overweight occurred in women.

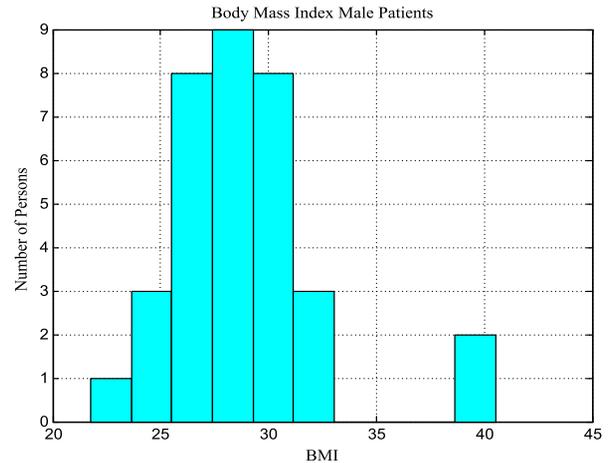


Figure 7 BMI of male patients
Source: Own [Matlab]

3. Clinical laboratory data

Different laboratory tests are performed daily from the first day to the last day of the patients' stay. There were 39 different laboratories that were obtained, although not all patients have sufficient data to be able to process them, so it was decided to perform a data cleaning in which it was reduced to 28 laboratories that are mentioned below: Glucose, Urea, Nitrogen Ureic, Creatinine, Total Proteins, Albumin, Globulin, A / G Ratio, Alanine Amino Transferase, Aspartate Amino Transferase, Total Bilirubin, Direct Bilirubin, Indirect Bilirubin, Lactic Dehydrogenase, Creatine-Phosphokinase, Chlorine, Potassium, Sodium, Time, Cytometry Prothrombin, Thromboplastin Time, Fibrinogen, Procalcitonin, Calcium, Phosphorus, Alkaline Phosphate, Magnesium, C. Reactive Protein (Pagana & Pagana, 2015)

4. Methodology

To carry out the statistical study of the CECCS data, a general block diagram was made, which is shown in Figure 8, each of the blocks has been described later. The data collection block corresponds to the collection of information contained in the hospital database, the information includes the graphs described in section 2, in addition to the data described in section 3. In blocks 2 and 3, the an analysis of the stored information, either to rule out errors, in addition to dividing the information into men and women, it is important to mention that several patients have missing data, for which the technique of imputation of the mean was used to add values missing in each laboratory of incomplete patients (Devore, 2012), (Gutierrez & Vladimirovna, 2014).

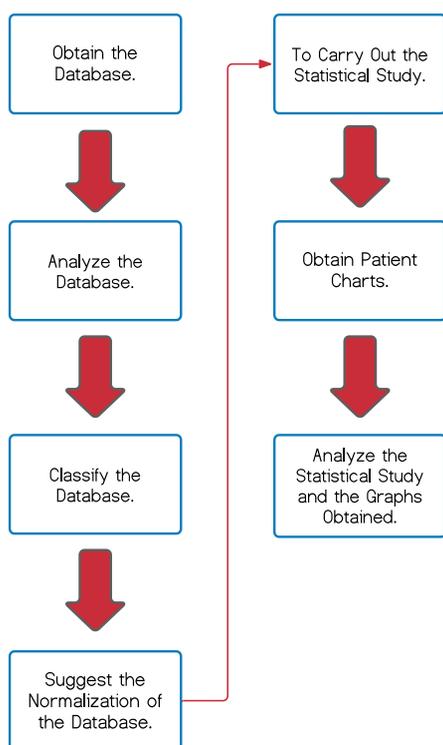


Figure 8 General block diagram
Own Source

Each one of the 28 laboratory studies in the database of each patient represents a vector of values, therefore, each vector is independent of each other, due to the above because each laboratory handles different representations measurement requires standardization of each of the data, given the above, a normalization from 0 to 1 was carried out, from the minimum and maximum values of each of the vectors independently (Gonzalez & Woods, 2008).

The fifth block is the realization of the statistical part, obtaining information on the minimum, maximum, mean, median values, the box and whisker diagrams, statistics of the behavior of the day-to-day laboratories and other extra statistics that were used Throughout the study, block six is only the display of statistics, while block seven is a study of the results obtained in the graphs and their subsequent interpretation (Devore, 2012), (Gutierrez & Vladimirovna, 2014).

5. Results

The statistical study and its interpretation will allow to know a little more about the virus and the behavior in the Guanajuato population, before which multiple graphs were made for both women and men.

Figure 9 represents the box and whisker diagrams of each of the laboratories of the women who were discharged from CECCS, the idea is to represent the averages in the changes in each laboratory, as well as their behavior during the different days. that lasted hospitalized.

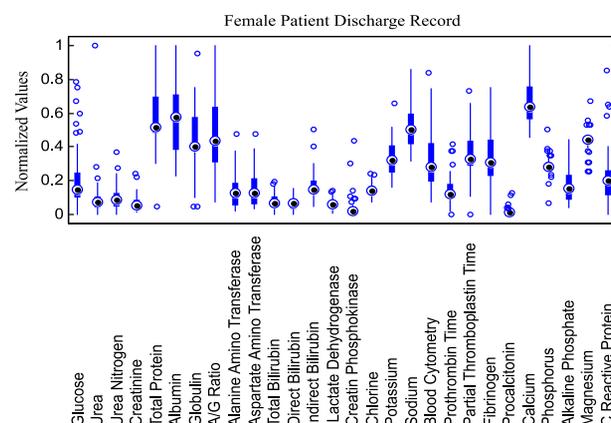


Figure 9 Laboratories of women registered
Source: Own [Matlab]

Each one of the graphs represents a sample of all the laboratories obtained from the women throughout the different days that they were admitted to the hospital.

In the case of Figure 10, the same graph is shown, but for female patients who died during their stay in the hospital, where multiple changes are observed compared to Figure 9, having higher mean values, higher variances, higher atypical values and in some cases very abrupt changes, such as in Glucose, Urea, Ureic Nitrogen and Total Proteins, as well as in the Partial Thromboplastin Time and C-Reactive Protein.

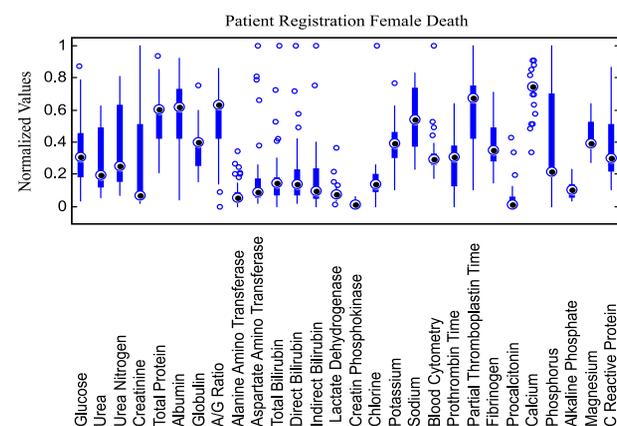


Figure 10 Laboratories of women who died
Source: Own [Matlab]

The graph in Figure 11 describes the case of the men who were discharged, where the diagrams of some laboratories with trends similar to those of the women who lived can be observed and their result is represented in Figure 9. Let us remember that the men's data is a sample of all the patients admitted to the CECCS, in which this sample of previously selected patients was saved in a database with all the laboratories obtained daily, to finally obtain the graphs of diagrams of boxes and whiskers shown.

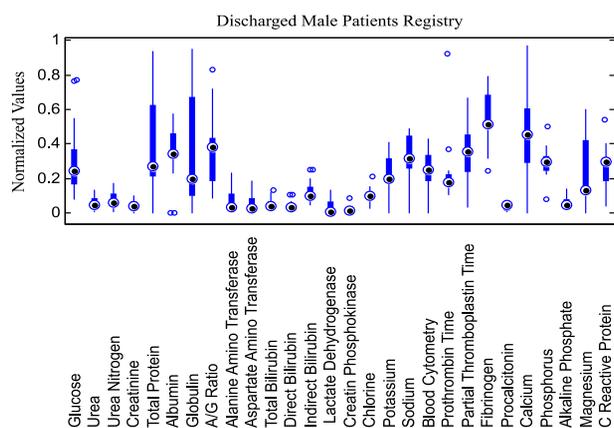


Figure 11 Labs of men discharged
Source: Own [Matlab]

Figure 12 shows the diagrams of the men who died, the graph shows greater changes in laboratories of C-Reactive Protein, Magnesium, Fibrinogen, Hematic Cytometry, Sodium, Glucose, Urea, Ureic Nitrogen, in addition to having more values atypical.

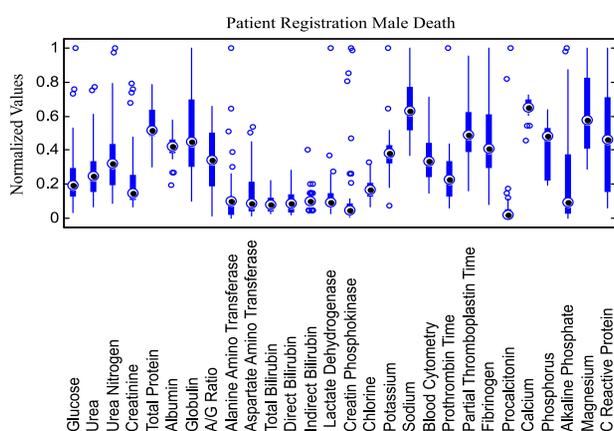


Figure 12 Laboratories of Men who died
Source: Own [Matlab]

From the previous graphs, certain relevant information can be observed, since, in the case of women, on average, they die with higher values in the different laboratories and also have wider variances, but it is not yet a strong conclusion. Therefore, more studies are required, before that, box and whisker diagrams were made to compare the laboratories of a man and a woman who lived and died.

Figure 13 describes the results of a woman, named as patient 14, who was discharged after 31 days, where lower outliers are observed and the variances are lower, as well as the means are lower.

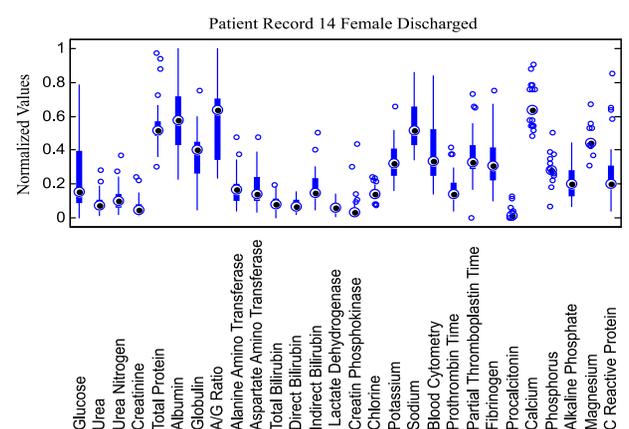


Figure 13 Laboratories for women discharged
Source: Own [Matlab]

Unlike the previous Figure, Figure 14 shows the results of patient 2, who died, in which there are more outliers, the mean is higher, as well as the variance in most laboratories.

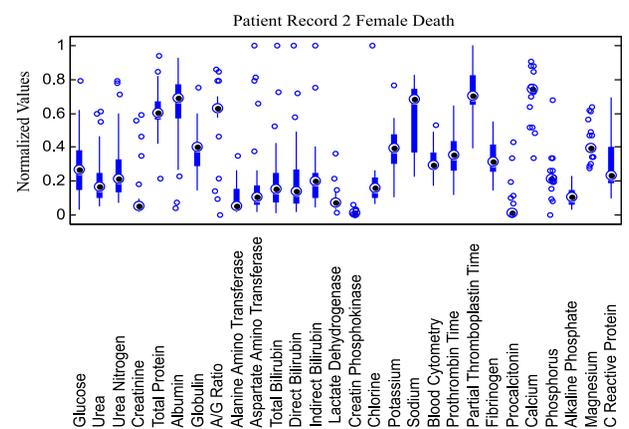


Figure 14 Laboratories of Woman who died
Source: Own [Matlab]

Figure 15 describes the results of a man, named as patient 16, who was discharged after 14 days, where lower outliers are observed and the variances are lower, as well as the means are lower.

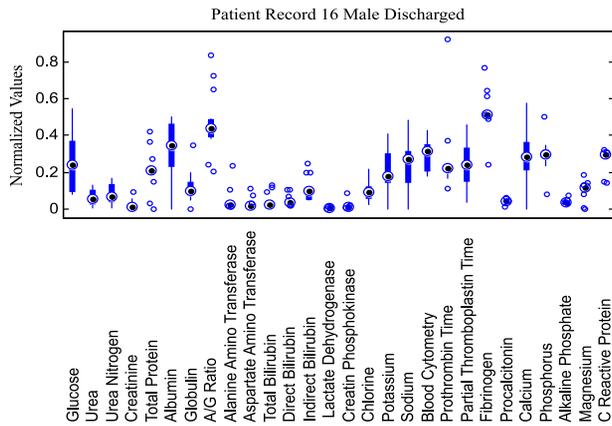


Figure 15 Laboratories of Man discharged
Source: Own [Matlab]

Compared to the previous Figure, Figure 16 shows the results of patient 17, who died after 13 days, in which there are more outliers, the mean is higher, as well as the variance in most laboratories, It is important to notice the strongest changes in Glucose, Urea, Ureic Nitrogen, Creatine-Phosphokinase, Sodium, Fibrinogen, Magnesium and C-Reactive Protein.

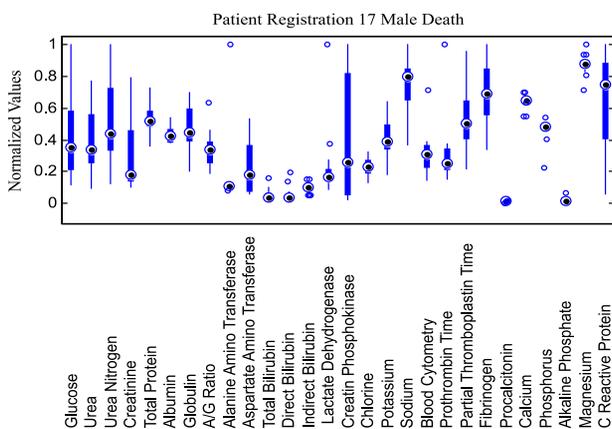


Figure 16 Laboratories of Man who died
Source: Own [Matlab]

The sample of 4 patients chosen to obtain their previous statistics describes the averages, variances, minimums and maximums of each laboratory. Likewise, graphs of the laboratory's day-to-day are added, in order to analyze the behavior of patients during their stay in the hospital.

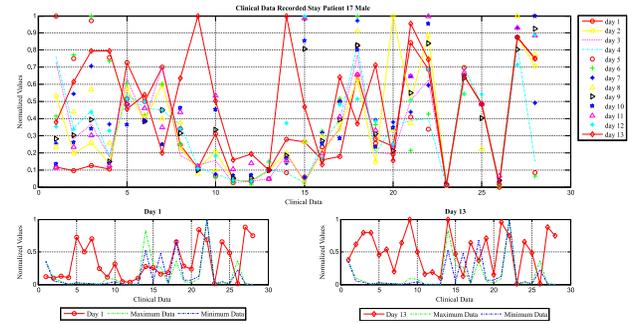


Figure 17 Patient behavior 17
Source: Own [Matlab]

The diagrams in Figure 17 show the laboratories obtained day by day during their stay, remembering that the patient died after 13 days, the Figure shows 3 graphs, the first one describes large variations in the values, the second graph (bottom left) represents the laboratory values of the first day in the hospital, likewise the minimum and maximum normal values of each laboratory are added, the third graph (lower right) shows the laboratory values of the last day of hospital stay, the values they are very high compared to the first day of stay.

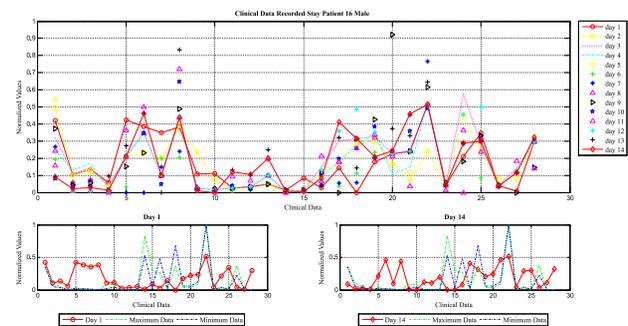


Figure 18 Patient behavior 16
Source: Own [Matlab]

For Figure 18 there are 3 graphs, the largest shows the behavior of patient 16 throughout the 14 days of stay, the lower left graph explains the first day of stay, the graph on the right side shows the last day of stay.

As in the previous figures, 3 graphs of patient 2 who died are shown, in Figure 19 it is possible to note the changes in values of each laboratory on the different days, showing the difference between the first day and the last day.

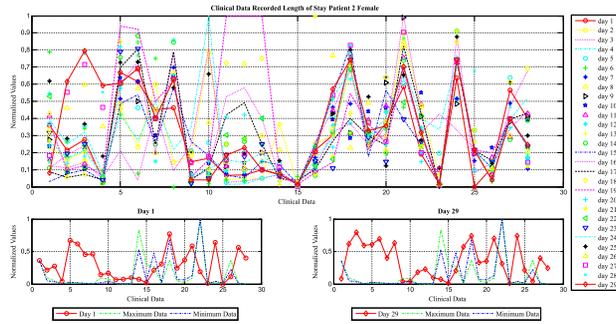


Figure 19 Behavior of patient 2
Source: Own [Matlab]

For patient 14 who was discharged after 31 days it is possible to analyze the laboratory results of the different days in Figure 20, then it shows a stabilization of the data at the end of her stay.

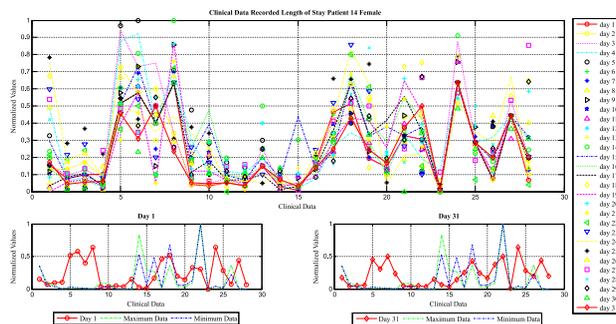


Figure 20 Behavior of patient 14
Source: Own [Matlab]

6. Conclusions

In this article, a statistical analysis of the patients admitted to the CECCS in Salamanca was carried out, which includes patients with COVID-19 and not COVID-19 from the state of Guanajuato, the graphs show trends similar to those observed at the national and international level.

The patients admitted were those with greater severity in the state, likewise, they were patients who were more overweight in both men and women, it is also important to mention that there were higher admissions in men than in women, in addition to the fact that more men were the who passed away.

From the sample of patients taken and derived from the statistics, it is also possible to analyze that women have greater resistance according to the graphs to COVID-19 than men and even have higher values in multiple laboratory results.

The statistics analyzed will make it possible to carry out a comparative study of the drugs supplied and their evolution during their stay, in addition to knowing the evolution of the patient, the use of ventilators and even making estimates of behavior or prevention of deaths in patients.

Acknowledgments

The authors would like to thank the State Center for Critical Care of Salamanca (CECCS), the doctor Dr. Cesar Centeno Fosado and the head of nurses Liliana Yañez Vargas for their support in preparing the article and providing the database.

References

Conversión del CECCS a centro COVID-19. (2020, 7 abril). Centro Estatal de Cuidados Críticos de Salamanca. <https://boletines.guanajuato.gob.mx/2020/04/07/ssg-reconvierte-centro-estatal-de-cuidados-criticos-en-unidad-especializada-para-atencion-de-pacientes-con-covid-19/>

Coronavirus – gov.mx. (2020, 14 diciembre). <https://coronavirus.gob.mx/>

Coronavirus disease (COVID-19). (2020, 14 diciembre). Coronavirus disease (COVID-19) pandemic. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>

Devore, J., 2012. Probability And Statistics For Engineering And The Sciences. 8th ed. Australia: Brooks/Cole, pp.28-38.

Dra. Celia M. Alpuche Aranda. (2020, junio). *Informe Técnico Reunión De Expertos Sobre Uso De Pruebas De Laboratorio Para Identificar SARS-COV-2 (N.º 1)*. Subsecretaría De Prevención Y Promoción De La Salud Dirección General De Epidemiología. https://coronavirus.gob.mx/wp-content/uploads/2020/06/Docto_InfTec-SARS-COV2_26jun2020.pdf

Gonzalez, R. C., & Woods, R. E. (2008). *Digital Image Processing* (3ª ed.). USA: Pearson/Prentice Hall.

Gutiérrez González, E. and Vladimirovna Panteleeva, O., 2014. Probabilidad Y Estadística. 1st ed. México: Grupo Editorial Patria, pp.287-323.

Liu, K., Fang, Y.-Y., Deng, Y., Liu, W., Wang, M.-F., Ma, J.-P., Xiao, W., Wang, Y.-N., Zhong, M.-H., Li, C.-H., Li, G.-C., & Liu, H.-G. (2020). Clinical characteristics of novel coronavirus cases in tertiary hospitals in Hubei Province. *Chinese Medical Journal*, 133(9), 1025-1031. <https://doi.org/10.1097/cm9.0000000000000744>

Liu, Y., Yang, Y., Zhang, C., Huang, F., Wang, F., Yuan, J., Wang, Z., Li, J., Li, J., Feng, C., Zhang, Z., Wang, L., Peng, L., Chen, L., Qin, Y., Zhao, D., Tan, S., Yin, L., Xu, J., ... Liu, L. (2020). Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. *Science China Life Sciences*, 63(3), 364-374. <https://doi.org/10.1007/s11427-020-1643-8>

Ng, M.-Y., Lee, E. Y. P., Yang, J., Yang, F., Li, X., Wang, H., Lui, M. M.-, Lo, C. S.-Y., Leung, B., Khong, P.-L., Hui, C. K.-M., Yuen, K.-, & Kuo, M. D. (2020). Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review. *Radiology: Cardiothoracic Imaging*, 2(1), e200034. <https://doi.org/10.1148/ryct.2020200034>

Pagana, K., & Pagana, T. (2015). *Laboratorio clínico: Indicaciones e interpretación de resultados* (1.^a ed.). Editorial El Manual Moderno.

Yun, H., Sun, Z., Wu, J., Tang, A., Hu, M., & Xiang, Z. (2020). Laboratory data analysis of novel coronavirus (COVID-19) screening in 2510 patients. *Clinica Chimica Acta*, 507, 94-97. <https://doi.org/10.1016/j.cca.2020.04.018>

Zhang, J.-, Dong, X., Cao, Y.-, Yuan, Y.-, Yang, Y.-, Yan, Y.-, Akdis, C. A., & Gao, Y.-. (2020). Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy*, 75(7), 1730-1741. <https://doi.org/10.1111/all.14238>