A Review of Chronic Diseases Map and Data Trends Analysis

Mustafa Yousef Mustafa ZGHOUL

Abstract — In this paper, an attempt is made to review and analysis the status of the chronic disease in the gulf region. In addition, illustrate the methods of visualization of the chronic diseases information, which are the major disease burden in the gulf region. This paper, is reviewed and discussed the methods of analysis the chronic diseases data with reference to univariate time series model. The Forecasting model based simple linear regression is proposed and implemented. The Visualization of digital data is considering as one of the most important techniques for presenting the distribution of statistical data within a small space. An interactive disease map is proposed and implemented to determine the chronic diseases numbers and locations.

Keywords: Chronic diseases, visualization, statistical data analysis, linear regression, time series, forecasting, mapping

I. INTRODUCTION

Chronic diseases (CD) pose a constant danger and a large force because of their increased risk to human life and the economy of nations. Statistics issued by the World Health Organization (WHO) indicate that the number of people living in chronic disease increasing in all parts of the world as illustrated in Figure 1.

Moreover, depending on the ministry of health statistics in the Sultanate of Oman during the period (1990 - 2005), approximately 75% of diseases burden is attributable to chronic diseases" [2]. The total number of death because of chronic diseases are a double number of death caused by infectious diseases, like pulmonary tuberculosis, viral hepatitis (A), malaria, and AIDS (HIV). In addition, the chronic diseases affect women and men at younger ages. There are a number of risk factors causes the chronic diseases including physical inactivity, unhealthy diet and use of tobacco. The chronic disease is a factual dangerous and it is growing rapidly over the time. Therefore, many governments spent billions of dollars for controlling and preventing the expansion of CD [3]. WHO reports indicate that the occurrence of CD will be raised dramatically in 2023 as shown in Figure 3.

For this reason, one of the solutions to reduce the rate of chronic diseases is increasing the health awareness in the community. Therefore, design and implement a web application for the chronic diseases surveillance is much needed, which will help in raising the knowledge of culture-related chronic diseases and methods of reducing its effects.
The chronic disease (or non-communicable disease) is a disease that remains or continues for a long period of time, starts from about three months or more. Generally, these diseases cannot be prevented using a treatment or take a medicine. Ministry of health in Sultanate of Oman is working on educating members of the community against the chronic diseases like diabetes, blood pressure, etc., in order to increase the health awareness. The infectious disease (or communicable disease) is a disease that generated by microorganisms, like bacteria, parasites, fungi and viruses. In addition, can be infected other people quickly through sneezing and coughing, or through physical contact. Examples of infectious diseases are malaria, pulmonary tuberculosis, viral hepatitis (A) and AIDS (HIV) [4]. Therefore, the government’s work hard to reduce the expansion of the chronic disease and infectious diseases by offering intensive programs and studies of injuries prevention methods [5]. The main objective of this paper is to promote awareness of the expansion of chronic diseases in GCC. Therefore, design and implement a web application for analyzing, forecasting and visualizing chronic diseases is much needed. This system will collect the data of chronic disease and illustrate the analysis statistics rates in all GCC. In addition, it will provide a very clear visualization figures based on interactive disease map.

II. VISUALIZATION METHODS AND TECHNIQUES

Many techniques are used to visualize and analysis the data. The interactive disease map is one of a powerful method for data visualization. It is used to illustrate information clearly and efficiently via plots, statistical and information graphics. The visualization of statistical data aims to present a large amount of information in the short time. Numerical data can be graphed using lines, bars or dots to visual communicate a quantitative message. Recent studies proved that the interactive disease map helps the users to analyze and understand the meaning of complex data easily. Generally, tables are used where the users look up to a specific measurement, while charts and maps are used to present patterns or relationships of multivariate data. The diseases map is a method for representing different diseases for tracking the expansion of diseases in real time and understanding the reasons of these diseases [6]. A time series is a group of observations or statistics which being recorded or collected at regular intervals and sorted by the time [7]. Therefore, the proposed system will provide an accurate analysis for predicting unseen data over time. Moreover, the trend analysis is a type of technical analysis that tries to predict the future values of data (information in sequence over time) based on past data [8],[9]. The analyze of the statistical data based on trends analysis method is called linear regression analysis and which will help the decision makers in the ministry of health to determine the expansion and needs of chronic diseases [10],[11].

Cartographic visualization is used symbolism techniques, which is referred to locations inside the map and represent their multiple data values [12]. In addition, it allows the users to select statistical and geographical subsets. Local statistics of univariate distributions can be calculated and visualized in a dynamic figure for exploration like scatterplots, dot-plots, population cartograms, choropleths, parallel coordinates plots, and polygon maps as shown in Figure 4.

According to Statistics Netherlands (2012), "data visualization is the art of presenting data in a visual manner. So the data becomes apparent. Data visualization is a helpful tool for all phases of the statistical process. It has two goals in statistics: data exploration and communication" [13]. Diagrams can detect the pattern in large amounts of data. This method views the major findings in the data to the users. There are many types of diagrams, like radar plot, bar chart, and line chart. It used to represent the time series or to make a comparison between two variables. On the other hand, proportional symbol map is a good way of viewing statistical data. It aims to view the characteristics of a subject on the map. "In a proportional symbol map, a symbol is plotted on the center of a region, and the surface area of the symbol is scaled with the value of the variable" [13]. For example, consider the proportional symbol map as shown in Figure 5.
According to Heinrich Hartmann (2016), "statistical techniques are the art of extracting information from data. Moreover, one of essential data analysis methods is visualization. The human brain can process geometric information much more rapidly than numbers." [14]. There are many methods for visualization, like rug plots, histograms (for one-dimensional data), scatter plots (for two-dimensional data), line plots. [14], as an example, consider the proportional symbol map as shown in Figure 6.

![Proportional symbol map source](image1)

**Fig. 5**: Proportional symbol map source [13]

According to Adriana REVEIU, Marian DARDALA (2011), cartographic visualization provides the facilities to represent statistical data. It is one of the most important tools in geographical information systems (GIS). In addition, it aims to show the distribution of statistical data inside the regional map. Moreover, it uses different symbols to show more information at the same time inside the map. These symbols view quantitative details for the users. [16]

III. ANALYSIS OF CHRONIC DISEASES

The analysis of chronic diseases often depends on longitudinal cohort (or group) studies, and there are many methodological issues related to chronic disease studies, the first one is based on changing definitions of risk factors and outcomes over time. Moreover, the second issue is missing data.

To perform analysis in the existence of missing data, there are many procedures to do that: the first one, make analysis to individuals which data is complete. The second one, insert the existing values to individuals which data is incomplete and then analyzing the dataset. Moreover, the second analytical technique is most appropriate. There are many analytic techniques for chronic diseases modeling: [17]

1. Logistic regression analysis: this can examine the effects of risk factors on the development of the disease.
2. Survival analysis (time to event data): this deal with time until the event of interest occurs.
4. Tree-based classification methods.
5. Longitudinal data analysis (mixed models, generalized linear models and generalized estimating equations).

The formula for representing the regression analysis model is:

\[ Y_i = a + b \times X_i \]

Where the regression parameters, \( a \) is the intercept (on the y axis) . And \( b \) is the slope of the regression line.

![Visualization methods source](image2)

**Fig. 6**: Visualization methods source [14]

![Visualization of varying data by symbol size on map source](image3)

**Fig. 7**: Visualization of varying data by symbol size on map source [15]
The Least Squares method used to estimate the slope and intercept regression parameters, as the following:

The formula for calculating slope regression parameter (b) is:

\[ b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \]

The formula for calculating intercepts regression parameter (a) is:

\[ a = \frac{\sum y}{n} - b \times \frac{\sum x}{n} \]

IV. LITERATURE SURVEY

This section presents the literature survey of the chronic diseases data analysis. The researchers were implemented different computing techniques for forecasting and visualizing the unseen data. It will cover the expansion rate of chronic diseases in the global. In addition, the forecasting models and visualization techniques are included. (WHO) global report [3] statistics shows that the number of deaths in 2005 is about (58) million, around (35) million approximately of deaths resulted by chronic diseases. It represents about 60% of global deaths, which caused by chronic diseases, "only 20% of chronic disease deaths occur in high-income countries, while 80% occur in low and middle-income countries, where most of the world’s population lives". The chronic diseases will rise about 70% of the total deaths in the world at 2030. Khatib O. [18] stated, "chronic diseases are the major disease burden in the Eastern Mediterranean Region. There are many risk factors associated with chronic diseases, most of them are related to the lifestyle and can be controlled. Such as low vegetable and fruit intake, physical inactivity, high fast food consumption and high cholesterol are dominant causes of cardiovascular disease and some types of cancer. Also obesity and overweight can raise the risk of chronic diseases, like heart disease and diabetes". There are many approaches can help to deal with this problem such as developing national strategies, policies, and plans for prevention and health care. Also, implement and enhance the community participation in prevention and health care. Most of the solutions for preventing chronic diseases are expensive [18]. Gregory Hartl & Menno van Hilten (2012) [1], explains that the non-communicable diseases (NCDs) are caused more than 60% of all deaths in the Gulf Cooperation Council (GCC) countries. The risk factors are an unhealthy diet, the use of tobacco and the physical inactivity. The Gulf Cooperation Council (GCC) countries (Saudi Arabia, Bahrain, Sultanate of Oman, Kuwait, Qatar, and United Arab Emirates) are adopting a regional strategy to address, prevent and control of non-communicable diseases (NCDs), like diabetes, cancer, and chronic respiratory disease. It aims to reduce exposure to people from different risk factors and improving the services of preventing and treating the health problems [1]. Hill A.G., et al. (200), mentioned that the awareness about the chronic diseases has been grown in Oman. Morbidity for diagnosis related to chronic diseases, especially cancer, cardiovascular disease, and the endocrine disease was growing and becoming a significant share of Oman’s burden of disease [19]. Jawad A. Al-Lawati et al. (2008), believes that the chronic diseases are posed the main challenge for Omani population [2]. Depending on the ministry of health statistics in Oman during the period (1990 to 2005), approximately 75% of diseases burden is attributable to chronic diseases. "The distribution of chronic diseases and related risk factors among the general population is similar to that of industrialized nations: 12% of the population has diabetes, 30% is overweight, 20% is obese, 41% has high cholesterol, and 21% has the metabolic syndrome". They conclude that the chronic diseases are the major exhaustion on human and financial resources for Sultanate of Oman, and this will affect the advances in the health care system that has been achieved. Similarly, some related works focus on using visualization techniques like interactive map. Jason Dykes (1998), implemented a cartographic visualization for locating symbols on a plane to show the statistical distributions of one or more variables” [12].
V. RESULTS AND DISCUSSIONS

Depends on statistical data which collected from ministry of health in sultanate of Oman, it shown that the prevalence of diabetes in Oman is increasing over years. as illustrated in Figure 8.

After applying the equation of simple linear regression, the equation will be \( Y = 4319.3 + 127.9 \times X \), and the value of R-squared variable equals to (0.46), this value will describe how data fitted to the regression line. As shown in Figure 9.

The results of this forecasting model shown that the prevalence of diabetes in sultanate of oman will increase. And this will help the decision makers to take attention about spread of chronic diseases.

This research paper will implement the simple linear regression method for statistical data analysis and forecasting because the collected data is one variable changed over time. In addition, this paper will select the cartographic visualization method to show the distribution of the data among different geographic locations. In addition, it uses symbols to represent the data; this symbol is scaled with the value of the variable.

CONCLUSION & FUTURE RESEARCH DIRECTION

This paper provided an overview of the chronic diseases in GCC. First, it presents the risk factors that causes the chronic diseases. Then it explains the different methods for statistical data analysis and forecasting models. Moreover, presents different methods for data visualization techniques is presented and reviewed. The Cartographic technique is appropriate for visualization of Chronic Diseases data because it illustrates the distribution of the data among different geographic locations. The literature survey proved that a simple linear regression is an appropriate method for data analysis and forecasting data with one variable only.

REFERENCES


[23] Christoph Klose, Marion Pircher, Stephan Sharma (May 2004), "Univariate Time Series Forecasting", UK.


[29] Peter Filzmoser, Karel Hron, Clemens Reimann (2009), "Univariate statistical analysis of environmental (compositional) data: Problems and possibilities", Austria, ScienceDirect, STOTEN:11466.
<table>
<thead>
<tr>
<th>S</th>
<th>Author</th>
<th>year</th>
<th>Place</th>
<th>Method of study</th>
<th>Major Findings</th>
<th>Merits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hill A.G., Muyeed A.Z., Al-Lawati J.A. [19]</td>
<td>2000</td>
<td>Oman</td>
<td>Numerical and Analytical</td>
<td>Diabetes has become a major chronic (or non-communicable) disease problem in Oman. The prevalence of diabetes varies among different governorates in Oman; the percentage for diabetes in 2000 is 13%.</td>
<td>The awareness about the chronic disease has been grown in Oman.</td>
<td>The percentage for diabetes in Oman was the highest reported in the arab countries.</td>
</tr>
<tr>
<td>2</td>
<td>Khatib O. [18]</td>
<td>2004</td>
<td>Oman</td>
<td>Analytical</td>
<td>The incidence of chronic diseases is rising in the middle east region. In 2000, 47% of the region’s load of disease is due to Non-communicable diseases and it is expected that this will rise up to 60% by the year 2020.</td>
<td>There are many strategies: developing a national strategies, policies and plans for prevention and care, also implement and enhance community participation in prevention and care.</td>
<td>The risk factors associated with chronic diseases are: physical inactivity, unhealthy diet and smoking.</td>
</tr>
<tr>
<td>3</td>
<td>World Health Organization [8]</td>
<td>2005</td>
<td>Geneva</td>
<td>Analytical and Numerical</td>
<td>The number of deaths in 2005 is about (58) million, around (35) million approximately of deaths resulted by chronic diseases. It represents about 60% of global deaths which caused by chronic diseases. If the current trends continue, chronic diseases by 2030 will rise about 70% of the total deaths globally.</td>
<td>Many governments around the world spent billions of dollars for medication to control and prevent the spread of chronic diseases</td>
<td>Chronic diseases are a threat and it is growing over the time and it will affect all countries.</td>
</tr>
<tr>
<td>4</td>
<td>Jawad A. Al-Lawati, Ruth Mabry</td>
<td>2008</td>
<td>Oman</td>
<td>Analytical</td>
<td>The chronic diseases pose the main challenge for</td>
<td>Health planners and decision makers in</td>
<td>Chronic diseases are the major exhaustion on</td>
</tr>
</tbody>
</table>
and Ali Jaffer Mohammed [20]

Omani population. During the period (1990 to 2005), approximately 75% of diseases burden is attributable to chronic diseases. Oman have greater commitment to the provision of services for people with chronic diseases. human and financial resources for Oman, and this will affect the advances in the health care system that has been achieved.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The chronic diseases cause more than 60% of all deaths in the Gulf Cooperation Council (GCC) countries. All the GCC countries adopting regional strategy to address prevent and control of chronic diseases.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The model for chronic disease consists of four stages or phases: disease free, pre-clinical (latent period), clinical manifestation, and follow-up. Good statistical approaches involve hypothesizing models for these stages, collecting appropriate data, and then fitting and testing the appropriate models. The risk factors are: age, gender, smoking status, blood pressure and cholesterol.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7</th>
<th>Kelly H. Zou, Kemal Tuncali, Stuart G. Silverman [27]</th>
<th>2003</th>
<th>USA</th>
<th>Simple Linear Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Simple linear regression measure the linear relationship between a predictor variable and an - The values of dependent variables can be estimated from the observed..</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- missing values is a common problem.</td>
</tr>
<tr>
<td>Page</td>
<td>Author(s)</td>
<td>Year</td>
<td>Country</td>
<td>Method/Model</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>------</td>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>8</td>
<td>Christoph Klose, Marion Pircher, Stephan Sharma [23]</td>
<td>2004</td>
<td>UK</td>
<td>Autoregressive moving average (ARMA) linear model</td>
</tr>
<tr>
<td>9</td>
<td>Peter Filzmoser, Karel Hron, Clemens Reimann [29]</td>
<td>2009</td>
<td>Austria</td>
<td>multivariate data analysis</td>
</tr>
<tr>
<td>10</td>
<td>Astrid Schneider, Gerhard Hommel, and Maria Blettner</td>
<td>2010</td>
<td>Germany</td>
<td>Linear Regression Analysis</td>
</tr>
<tr>
<td>Reference</td>
<td>Authors</td>
<td>Year</td>
<td>Country</td>
<td>Methods</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>[22]</td>
<td>Choong-Yeun Liong and Sin-Fan Foo</td>
<td>2013</td>
<td>Malaysia</td>
<td>logistic regression (LR) analysis and Linear discriminant analysis (LDA)</td>
</tr>
<tr>
<td>11</td>
<td>Adela SASU</td>
<td>2013</td>
<td>Romania</td>
<td>ARIMA, Linear Regression (LR), Multilayer perceptrons network (MLP)</td>
</tr>
<tr>
<td>13</td>
<td>Suzilah Ismaila, Rohaiza Zakariaa</td>
<td>2014</td>
<td>Malaysia</td>
<td>Quantitative (Projective)</td>
</tr>
<tr>
<td>ID</td>
<td>Author(s)</td>
<td>Year</td>
<td>Country</td>
<td>Technique</td>
</tr>
<tr>
<td>----</td>
<td>---------------------------------</td>
<td>------</td>
<td>---------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>14</td>
<td>Jason Dykes [12]</td>
<td>1998</td>
<td>UK</td>
<td>Cartographic visualization</td>
</tr>
<tr>
<td>15</td>
<td>Ben Fry [15]</td>
<td>2008</td>
<td>USA</td>
<td>Mapping</td>
</tr>
<tr>
<td>16</td>
<td>Adriana REVEIU, Marian DARDALA</td>
<td>2011</td>
<td>Romania</td>
<td>Cartographic visualization</td>
</tr>
</tbody>
</table>
| 17 | Statistics Netherlands [13] | 2012 | Netherla nds | Proportional symbol map | Proportional symbol map is a good way of viewing statistical data. It aims to view the characteristics of a subject on the map. | - present the geographic location and the distribution of the data.  
- make comparison between different areas.  
- summarise large amounts of data.  
- It is difficult to calculate the actual value (if not shown).  
- It consumes more time to execute.  
- The size of symbol may hide the location on map. |
|---|---|---|---|---|---|
| 18 | Mohammad S. Alam [28] | 2013 | Russia | - Tree-map  
- Circle Packing  
- Sunburst  
- Circular Network Diagram | One of the most requirements that the analyst tools should meet is to show more than one view per representation display. Circular Network Diagram is the most appropriate method for visualizing the statistical data. | - Tree-map: hierarchical grouping clearly shows data relations.  
- Circle Packing: space-efficient.  
- Sunburst: easily perceptible by most humans.  
- Circular Network Diagram: allows us to make relative data representation. And inside the circle, the resolution varies linearly, increasing with radial position.  
- Tree-map: not suitable for examining historical trends and time patterns.  
- Circle packing: same as for Tree-map method.  
- Sunburst: same as for Tree-map method.  
- Circular Network Diagram: objects with the smallest parameter weight can be suppressed by larger ones. |
| 19 | Heinrich Hartmann [14] | 2016 | USA | Histograms  
Line plots | One of the most essential data analysis methods is visualization. The Histogram is a popular visualization | - Histogram view number of values within interval.  
- Histogram used only for numerical.  
- Line plot is difficult to...
<table>
<thead>
<tr>
<th><strong>Rug plots</strong></th>
<th><strong>Scatter plot</strong></th>
<th>method. It is commonly used to show the distribution of numerical data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line plot show changes in data over time.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scatter plot show the relationship between two variables.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>read accurately, if there is a wide range of data.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scatter plot cannot show the relation of more than two variables.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>