

Statistical Modelling for Prediction of Anemia Among Women at Reproductive Age.

Rutuja Taware, Ajinkya Date, Ankita Jadhav, Vishal Jadhav.

Department of Statistics, Tuljaram Chaturchand College Baramati, Pune.

Abstract:

The research is investigated prediction of anaemia among women at reproductive age in Baramati. Iron deficiency is major cause of anaemia and is more prevalent in developing countries, posing additional burden on health care systems, in the presence of scarcity of resources. The high-risk group consists of females due to physiological reasons. Multiple factors can predict anaemia, which in turn leads to various adverse outcomes.

Iron deficiency is a major cause of anemia and is more prevalent in developing countries, posing additional burden on health care systems, in the presence of scarcity of resources. The high-risk group consists of females and children but comparatively it is higher in females due to physiological reasons. Multiple factors can predict anemia, which in turn leads to various adverse outco

Iron deficiency is a major cause of anemia and is more prevalent in developing countries, posing additional burden on health care systems, in the presence of scarcity of resources. The high-risk group consists of females and children but comparatively it is higher in females due to physiological reasons. Multiple factors can predict anemia, which in turn leads to various adverse outco

Iron deficiency is a major cause of anemia and is more prevalent in developing countries, posing additional burden on health care systems, in the presence of scarcity of resources. The high-risk group consists of females and children but comparatively it is higher in females due to physiological reasons. Multiple factors can predict anemia, which in turn leads to various adverse outco

ckground: Iron deficiency is a major cause of anemia and is more prevalent in developing countries, posing additional burden on health care systems, in the presence of scarcity of resources. The high-risk group consists of females and children but comparatively it is higher in females due to physiological reasons. Multiple factors can predict anemia, which in turn leads to various adverse outcom

Aim:

1. To explore the age, Haemoglobin, Red blood cells (R.B.C), Mean corpuscular volume, associated with Anaemia among the Shantabai Deshpande Memorial Hospital women's and investigate the relationship between Haemoglobin, Red blood cells, Mean corpuscular volume.

2. To determine Anaemia from Haematology.

Methods:

The study followed a cross-sectional study design. A collection of data from Deshpande Hospital Baramati. The project is focused on medical and health evaluation of anaemia. To determine anaemia from haematology is the goal research. Haematology is the study of blood and blood disorders.

Keywords:

Age, Haemoglobin, Red blood cells (RBC), Packed cell volume (PCV), mean corpuscular volume (MCV), mean corpuscular Haemoglobin (MCHC), Mean corpuscular Haemoglobin concentration (MCHC.Con), red cell distribution width (RDW).

Date of Submission: 06-08-2022

Date of Acceptance: 21-08-2022

I. Introduction

Iron deficiency anaemia arises when the balance of iron intake, iron stores and the body's loss of iron are insufficient to fully support production of erythrocytes. Iron deficiency anaemia rarely causes death, but the impact on human health is significant. In the developed world, this disease is easily identified and treated, but frequently overlooked by Data Analyst. In contrast, it is a health problem that affects major portions of the population. Overall, the prevention and successful treatment for iron deficiency anaemia remains woefully insufficient worldwide, especially among underprivileged women. Here, clinical and laboratory features of the disease are discussed, and then focus is placed on reproductive age group (15-49) among global populations. Anaemia is condition in which you lack enough healthy red blood cells to carry adequate oxygen to your body tissues. Having anaemia can make you feel tired and weak.

There are many forms of anaemia such as Mild, Moderate, Severe, No anaemia. Each with its own cause anaemia can be temporary or long term, and it can certain recommended level. Mild deficiency consists of depleted stores with normal production of Hb and iron-dependent protein. Moderate deficiency consists of depleted iron stores and decreased iron-dependent protein production but normal Hb concentration. Severe anaemia consists of decreased concentration of circulating RBCs which results in decreased concentration of Hb within blood cells resulting in compromised transport of oxygen to tissues, iron stored are further depleted and the concentration of iron-dependent oxidative enzymes are reduced. According to WHO the recommended haemoglobin level, for women reproductive age group (WRA) (15-49 years) is 120 gm/L most blood cells, including red blood cells, your body needs iron, vitamin B-12, folate and other nutrients from the food you eat.

Recognizing Iron Deficiency Anaemia- Haemoglobin values used for the definition of anaemia vary according to Age, sex and weather the blood sample from the capillary or venous source. In the absence of automated testing, portable devices or visual matching of haemoglobin colour have been proven to be 95% accurate for identifying the haemoglobin level within 1-2g/dL of reference values.

Objective

- To identify the type/ stage of anaemia.
- To predict anaemia from haematology.
- To determine prevalence of anaemia according to age.

II. Material and Methods

This project undergoes with primary data collection through Medical Foundation & Research Centre Pvt. Ltd. The population was divided into age group. Dependent variable is (Haemoglobin). Blood samples for the Haemoglobin test were collected from women who voluntarily provided their consent to undertake the test and otherwise excluded. According to WHO, for reproductive age of women defined as blood Haemoglobin level <12.0 g/dL, which was further categorized as mild (11-1111.9g/dL), moderate(8.0-10.9 g/dL), severe anaemia(<8.0 g/dL), no anaemia(>12.0g/dL). For analysis, the severity of anaemia was categorized as any-anaemia and no- anaemia.

The independent variable in this study included The Age(15-24 years, 25.34 years, 35-49years) Red blood cells, Packed cell volume, mean corpuscular volume, mean corpuscular Haemoglobin, Mean corpuscularHaemoglobin concentration, red cell distribution width.

Study design

This cross-sectional study was carried out at the Shantabai Deshpande Memorial Institute of critical care medicine (Department of pathology) Baramati, Pune. Collection of data from hospital by using CBC count report. Types of studies included were exploratory data analysis (EDA) studies, correlational studies, decision tree, logistic regression, K-Nearest neighbours (KNN).

III. Methodology

The project undergoes with primary data collection through medical foundation & research centre Pvt. Ltd. The data was collected from n=345 women and 8 variables such as age, Hb, Red blood cells (RBC), Packed cell volume (PCV), mean corpuscular volume (MCV), mean corpuscular Haemoglobin (MCH), Mean corpuscular Haemoglobin concentration (MCHC), red cell distribution width (RDW-CV). The dependent variable is haemoglobin (Hb) and other seven are independent variable. This project focused on the variable which leads to anaemia in women's reproductive age. We also taken reference from Dr. Mrs. Madhuri Raut for this project.

Exploratory data analysis (EDA)concludes that the most of the women belongs to MCV.From entropy the most significance variable MCHC and PCV.We convert our data into binary data for performing logistic regression in order to find model adequacy and significant variable based on p-value of the deviance. Used machine learning algorithm like KNN, Logistic regression, Decision tree, to find accuracy of model. KNN

shows 86.7% accuracy, logistic regression shows 77.5% accuracy & decision tree shows 94.1% accuracy for the data. Age Groupwise prevalence of Anaemia was concluded by using Multiple bar graph. From multiple bar graph we conclude that most of the women n=104 belongs to age group 15-25 years. The percentage of mid anaemia on age group 15-25 year is highest.

Population

Women's in Shantabai Deshpande Memorial Institute. Baramati, pune

Inclusion Criteria

- Age group between 15-49 years (Reproductive age)
- Various Women's who are tested there CBC count.

Exclusion criteria

- Age not less than 15 years.

Statistical Analysis

Data were explored by using Exploratory data analysis(EDA). Analysis was performed on python (jupyter notebook), IBM SPSS Statistics, R programming, Excel. For the research statistical design used like machine learning algorithms, regression analysis.

IV. Discussion

We aimed to predict anaemia from haematology, to determine prevalence of anaemia according to age, to identify the type/stage of anaemia. PCV and MCHC are the most significance variable as per entropy. Packed cell volume it is ratio of plasma to RBC. The range of PCV is between 40-50% PCV shows low entropy that means all women's have normal PCV. MCHC is the average amount of concentration of haemoglobin per cell. MCHC never be high than requirement if it should low then it associate with peptic ulcer, malabsorption, iron deficiency. Among the proximal determinants, an increase in healthcare utilization, including antenatal care and iron supplementation is most important. Because of anaemia women suffer with fatigue, vitamin B12 deficiency, certain medicines, Folate deficiency, lose blood during menstruation, blood loss during childbirth, infants. This are the main anaemia causes which effects on women's health.

Exploratory data analysis analysing main characteristics which related with our research. From dependent variable(Hb) we conclude that average haemoglobin of all women's is up to 11.33. Required percentage for normal Hb is 11.6-15g/dL. Improvements in dietary diversity should eat animal source food by its content in vitamin B12. From this research 2.3% women shows severe level of anaemia. Anaemia can lead to rapid or irregular heartbeat (arrhythmia). When you are anaemic your heart pumps more blood to make up for the lack of oxygen in the blood. This can lead to enlarged heart or heart failure, death. 29.3% women shows moderate level of anaemia, 33.6% women are mild and 34.8% shows no anaemia. To avoid stages of severe anaemia we should take iron supplement, take a vitamin C rich fruit dessert this vitamin increases the absorption of iron from food, transfusions of red blood cells.

On other hand, the relationship between Hb and mean corpuscular volume is also observed in settings where the burden of disease increased the Macrocytic anaemia leads and when decrease MCV Microcytic anaemia will occurs. According to prevalence women belongs from age group 15-25 have mild anaemia is highest. Improved dietary intake, especially an increase in consumption of iron rich foods, folic acid helps to from RBCs, also predicted improvements in WRA anaemia prevalence.

V. Limitation:

Since we only included studies using observational data to assess the trends in WRA anaemia prevalence and its determinants and drivers, it is not possible to infer causality. We also only found three studies that carried out regression decomposition analysis, considered to be one of the most comprehensive and rigorous statistical methods for analysing observational data. However, the fact that several common determinants emerged, across the countries, geographies.

VI. Results:

Among 345 women's 2.3% women shows severe level of anaemia, 29.3% women shows moderate level anaemia, 33.6% women shows mild and 34.8% women shows no anaemia. Decision tree gives 94% the best accuracy for the dataset. From decision tree Mean corpuscular Hb concentration and Packed cell volume shows low entropy and that are most significance variable.

Age groupwise prevalence of Anaemia conclude that most of the women belongs in age group 15-25. The percentage of mild anaemia in age group 15-25 is highest (n=104) women.

Logistic regression conclude that fitted model is adequate. From odds ratio PCV & MCV have odds ratio greater than 1 which shows the best variables for model. From ANOVA RBC count, MCH, MCHC are the significant variables.

Rutuja Taware, et. al. "Statistical Modelling for Prediction of Anemia Among Women at Reproductive Age." *IOSR Journal of Dental and Medical Sciences (IOSR-JDMS)*, 21(08), 2022, pp. 07-10.