Bimodal Distribution of Blood Glucose and Non-Glucose Markers of Diabetes: A Statistical Study

Jyoti Thanvi, M.Sc., Ph.D.[1]
Shailendra Paliwal, B.Tech[2]
Krishna Kumar Sharma, M.Pharm[3]
Rajeev Gupta, M.D., Ph.D.[3]

From: *[1]Department of Statistics, Indira Gandhi National Tribal University, Amarkantak, MP, [2]Department of Mathematics and Statistics, Jaipur Engineering College and Research Centre, Jaipur, India; [2]Department of Preventive Cardiology and Internal Medicine, Eternal Heart Care Centre and Research institute, Mount Sinai New York Affiliate, Jaipur, India.*

Correspondence:*Dr. Jyoti Thanvi, [1]Department of Statistics, Indira Gandhi National Tribal University, Amarkantak, MP*

ABSTRACT

Objective: *We hypothesized that there is bimodal distribution of fasting blood glucose and others markers of diabetes (waist circumference, waist-hip ratio, triglycerides, HDL cholesterol) in India. To test this hypothesis we studied adult participants in India Heart Watch study where a high prevalence of diabetes has been reported.*

Methods: *A study was conducted in 11 cities in different regions of India using cluster sampling. Participants were evaluated for demographic, biophysical, and biochemical risk factors. 6198 participants were recruited, and in 5359 participants (86.4%, men 55%), details of diabetes (known or fasting glucose >126 mg/dL), hypertension (known or blood pressure>140/>90 mm Hg), hypercholesterolemia (cholesterol >200 mg/dL), low high-density lipoprotein (HDL) cholesterol (men <40, women <50 mg/dL), hypertriglyceridemia (>150 mg/dL), and smoking/tobacco use were available. Details of awareness, treatment, and control of hypertension and hypercholesterolemia were also obtained.*

Conclusions: *By plotting kernel density graphs and using Hartigans' dip test we can conclude that all most all the major risk factors i.e. weight, waist-circumference, waist hip ratio and fasting glucose etc. are having bimodal distribution in the study.*

INTRODUCTION

Most cardiovascular risk factors follow a normal statistical distribution.[1] Studies have reported that all the major risk factors- height, weight, body mass index, systolic and diastolic blood pressure, cholesterol, LDL cholesterol- when reported as numerical variables have a bell shaped frequency distribution graph. This distribution also extends to lifestyle behaviors- smoking, alcohol abuse, unhealthy diet, physical inactivity, stress and mental diseases.[1] This distribution has a great significance for prevention strategies. Geoffrey Rose in his seminal book - *The Strategy of Preventive Medicine* - suggested that population-wide prevention strategies are useful because these efforts lead to a rightward shift of the normal distribution with resultant lowering of absolute risk.[2]

There has been a sea-change in cardiovascular risk factors in the last twenty years. Instead of smoking-hypertension concept of cardiovascular risk, the new paradigm involves importance of cardiometabolic risk with dyslipidemia, diabetes and metabolic syndrome as major drivers of worldwide cardiovascular epidemic.[3] Studies have reported that most cardiovascular risk factors including fasting blood glucose follows a normal unimodal distribution, especially in Caucasian populations.[4] Studies among the non-Caucasian populations as well as American Indians have suggested a bimodal distribution of blood glucose levels.[5-10] It has been suggested that populations that have a high prevalence of diabetes have a bimodal distribution although these studies are limited to study of American Indians and Hispanic populations in USA and Pacific Islanders. [11] Recent studies regarding usefulness of bimodal distribution of blood fasting and 2-h post-meal glucose in diagnosis of diabetes have questioned this approach.[10,11] Moreover, if fasting and 2-hour postprandial blood glucose as well as HbA1c levels follows a bimodal distribution, as reported in NHANES study,[10] it is likely that early markers of diabetes such as waist circumference, waist-hip ratio, serum triglycerides and HDL cholesterol should have a likewise distribution. Diabetes is highly prevalent in Asia, especially in China, India and West Asia.[12] Population distribution of blood glucose levels has not been reported from these countries. Statistical distribution of phenotypic markers of diabetes and impaired glucose tolerance (waist circumference, waist-hip ratio, serum triglycerides, HDL cholesterol etc.) have also not been studied. We hypothesized that there is bimodal distribution of fasting blood glucose and others markers of diabetes (waist circumference, waist-hip ratio, triglycerides, HDL cholesterol) in India. To test this hypothesis we studied adult participants in India Heart Watch study where a high prevalence of diabetes has been reported. [13]

METHODS

A multisite study to identify prevalence of cardiovascular risk factors and their socio-demographic determinants was organised among urban subjects in India. Rationale and detailed methodology of this study has been reported earlier.[13,14] Protocol was approved by the institutional ethics committee of the national coordinating centre. Written informed consent was obtained from each participant. The study proforma was developed according to recommendations of the World Health Organization (WHO).[15]

We planned the study to identify prevalence of cardiometabolic risk factors and their determinants among urban subjects in India. Medium sized cities were identified in each of the large states of India and investigators who had a track record of research in cardiovascular or diabetes epidemiology were invited to participate. 20 investigators were invited from all large states of India and 11 participated. The cities are in northern (Jammu, Chandigarh, Bikaner), western (Ahmadabad, Jaipur), eastern (Lucknow, Patna, Dibrugarh) and southern (Madurai, Belgaum, Nagpur) regions of India.[16] The study data were collected in the years 2006-2010. Simple cluster sampling was performed at each site. A middle-class location was identified at each city based on municipal classification derived from cost of land, type of housing, public facilities (roads, sanitation, water supply, electricity, gas supply, etc.), and educational and medical facilities as reported earlier. 16 Sample size of about 250 men and 250 women (n=500) at each site is considered adequate by the WHO to identify 20% difference in mean level of biophysical and biochemical risk factors.[15] Accordingly, we invited 800-1000 subjects in each location to ensure participation of at least 500 subjects at each site estimating a response of 70%. The surveys were preceded by meetings with community leaders to ensure good participation. Subjects were invited in fasting state to a community centre of medical centre within each locality either twice or thrice a week depending upon the investigator's schedule. Inclusion criteria were all adults aged >20-75 years living in the particular location. Subjects who were confined to home with severe debilitating disease, those not likely to survive beyond 6 months and pregnant women were excluded.

The research worker employed by the site investigator after details were inquired from the subject filled the study form. Apart from demographic history, details of educational status, history of known hypertension, diabetes, lipid abnormalities and cardiovascular disease were inquired. Details regarding smoking and smokeless tobacco use, alcohol intake, dietary fat and fruits and vegetables intake as well as physical activity were assessed. Equipment for measurement of height, weight, waist and hip size and blood pressure were similar to ensure uniformity.[15] Sitting blood pressure was measured after at least 5 minute rest using standardized instruments. Three readings were obtained and were averaged for the data analysis. Fasting blood sample was obtained from all individuals after 8-10 hours fast. Fasting state was determined according to self-reports. The blood samples were obtained at community centers by technicians from an accredited national laboratory - Thyrocare Technologies Ltd., Mumbai, India (www.thyrocare.com). Blood glucose was measured at the local biochemistry facility of these laboratories. Blood for cholesterol, cholesterol lipoproteins and triglycerides estimation was transported under dry ice to the national referral laboratory where all the blood samples were analyzed using uniform protocol. Cholesterol, high density lipoprotein (HDL) cholesterol and triglyceride levels were measured using enzyme-based assays with internal and external quality control (www.thyrocare.com). Values of low density lipoprotein (LDL) cholesterol were calculated using Friedewald's formula (LDL cholesterol = Total cholesterol - HDL cholesterol + (triglycerides/5)).[17]

Statistical analyses: All the data was analyzed and plots were generated using R version 3.1.2 (2014-10-31)[18] and using R packages.[19] More than 90% data for various variables were available and in more than 85% subjects the data for all the variables were available. Diabetes prevalence was identified by presence of known history of diabetes or fasting blood glucose >126 mg/dl. We calculated descriptive statistics and plotted distribution of fasting blood glucose and other markers of diabetes- waist size, waist-hip ratio, waist-height ratio and blood triglycerides. Kernel density graphs were created and as per suggested by the graph we used the Hartigans' dip test to confirm our hypothesis.

RESULTS

Written informed consent was obtained from each participant. The study proforma was developed according to recommendations of the World Health Organization (WHO).[15]
We planned the study to identify prevalence of cardiometabolic risk factors and their determinants among urban subjects in India. Medium sized cities were identified in each (Madurai, Belgaum, Nagpur) regions of India.[16] The study data were collected in the years 2006-2010. Simple cluster sampling was performed at each site. A middle-class location was identified at each city based on municipal classification derived from cost of land, type of housing, public facilities (roads, sanitation, water supply, electricity, gas supply, etc.), and educational and medical facilities as reported earlier.[16] Sample size of about 250 men and 250 women (n=500) at each site is considered adequate by the WHO to identify 20% difference in mean level of biophysical and biochemical risk factors.[15] Accordingly, we invited 800-1000 subjects in each location to ensure participation of at least 500 subjects at each site estimating a response of 70%. The surveys were preceded by meetings with community leaders to ensure good participation. Subjects were invited in fasting state to a community centre of medical centre within each locality either twice or thrice a week depending upon the investigator's schedule. Inclusion criteria were all adults aged >20-75 years living in the particular location. Subjects who were confined to home with severe debilitating disease, those not likely to survive beyond 6 months and pregnant women were excluded.

The research worker employed by the site investigator after details were inquired from the subject filled the study form. Apart from demographic history, details of educational status, history of known hypertension, diabetes, lipid abnormalities and cardiovascular disease were inquired. Details regarding smoking and smokeless tobacco use, alcohol intake, dietary fat and fruits and vegetables intake as well as physical activity were assessed. Equipment for measurement of height, weight, waist and hip size and blood pressure were similar to ensure uniformity.[15] Sitting blood pressure was measured after at least 5 minute rest using standardized instruments. Three readings were obtained and were averaged for the data analysis. Fasting blood sample was obtained from all individuals after 8-10 hours fast. Fasting state was determined according to self-reports. The blood samples were obtained at community centers by technicians from an accredited national laboratory- Thyrocare Technologies Ltd., Mumbai, India (www.thyrocare.com). Blood glucose was measured at the local biochemistry facility of these laboratories. Blood for cholesterol, cholesterol lipoproteins

and triglycerides estimation was transported under dry ice to the national referral laboratory where all the blood samples were analyzed using uniform protocol. Cholesterol, high density lipoprotein (HDL) cholesterol and triglyceride levels were measured using enzyme-based assays with internal and external quality control (www.thyrocare.com). Values of low density lipoprotein (LDL) cholesterol were calculated using Friedewald's formula (LDL cholesterol = Total cholesterol - HDL cholesterol + (triglycerides/5)).[17]

Statistical analyses: All the data was analyzed and plots were generated using R version 3.1.2 (2014-10-31)[18] and using R package, diptest[19]. More than 90% data for various variables were available and in more than 85% subjects the data for all the variables were available. Diabetes prevalence was identified by presence of known history of diabetes or fasting blood glucose >126 mg/dl. We calculated descriptive statistics and plotted distribution of fasting blood glucose and other markers of diabetes- waist size, waist-hip ratio, waist-height ratio and blood triglycerides. Kernel density graphs were

The study was performed at 11 cities in India. 6198 subjects (men 3426, women 2772) of the targeted 9,900 subjects were evaluated (response 62%). Data for social and demographic characteristics and various cardiovascular risk factors have been reported.[13,14,16,17] Age-group specific prevalence of diabetes and impaired fasting glucose with 95% confidence intervals is shown in Table 1. There is a significantly increasing trend in diabetes as well as impaired fasting glucose in both men and women with increasing age. Impaired fasting glucose is greater than diabetes at age less than 40 years and is lower thereafter (Figure 1).

Descriptive statistics of blood glucose and other markers of diabetes body mass index, waist circumference, waist-hip ratio, waist-height ratio and serum triglycerides in men and women are shown in Tables 2 and 3. There is significant skewness in distribution of fasting blood glucose, waist circumference, waist-hip ratio and triglycerides in both men and women. Non-proximate markers of diabetes such as height, weight, body mass index and waist-height ratio do not show similar skewness.

We plotted distribution graphs of blood glucose in men and women at various age-groups (<35y, 35-49y and 50 years+) and superimposed with a density curve with Gaussian kernel. The red line shows the mean of specific group (Figure 1). There is a positive skewness in the distribution which progressively increases with a clear bimodal distribution at age >50 years. We also plotted other markers of diabetes (waist size, waist-hip ratio and blood triglycerides) in men and women in the study cohort. The graphs (Figures 2, 3 and 4) demonstrate a skewed bimodal distribution of these variables. On the other hand variables that are not a strong marker of diabetes, e.g., body-mass index, shows a normal distribution (Figure 5).

With the help of Hartigans' dip test we confirmed the bimodality of variables in the population under study. Major variables which weight, waist circumference, waist-hip ratio, fasting glucose etc., but bimodality of triglyceride was not confirmed and normality of BMI was confirmed by the test.

REFERENCES

1. *Bhopal R. Concepts of epidemiology: An integrated introduction to the ideas, theories, principles and methods of epidemiology. Oxford. Oxford University Press. 2002; 1-42.*
2. *Rose G. The strategy of preventive medicine. Oxford. Oxford University Press. 1992; 64-94.*
3. *Benjamin EJ, Smith SC, Cooper RS, Hill MN, Luepkar RV. Task Force #1: Magnitude of the prevention problem, opportunities and challenges. JAm CollCardiol. 2002; 40:588-603.*
4. *Rose G, Day S. The population mean predicts the number of deviant individuals. BMJ. 1990; 301;1031-1034.*
5. *Rushforth NB, Bennett PH, Steinberg AG, Burch TA, Miller M. Diabetes in Pima Indians: evidence of bimodality in glucose tolerance distribution. Diabetes. 1971; 20:756-765.*
6. *Zimmet P, Whitehouse S. Bimodality of fasting and 2-hour glucose tolerance distributions in a Micronesian population. Diabetes. 1978; 27:793-800.*
7. *Rosenthal M, McMahan CA, Stern MP, Eifler CW, Haffner SM, Hazuda HP, Franco LJ. Evidence of bimodality of two hour plasma glucose concentrations in Mexican Americans: results from the San Antonio Heart study. J Chronic Dis. 1985; 38:5-16.*
8. *Lim TO, Bakri R, Morad Z, Hamid MA. Bimodality in blood glucose distribution: is it universal? Diabetes Care. 2002; 25:2212-2217.*
9. *Fan J, May SJ, Zhou Y, Barrett-Connor E. Bimodality of 2-h plasma glucose distribution in Whites: the Rancho Bernardo study. Diabetes Care. 2005; 28:1451-1456.*
10. *Vistisen D, Colaguiri S, Borch-Johnson K, on behalf of the DETECT-2 Collaboration. Bimodal distribution of glucose is not universally useful for diagnosing diabetes. Diabetes Care. 2009; 32:397-403.*
11. *Menke A, Rust KF, Savage PJ, Cowie CC. Hemoglobin A1c, fasting plasma glucose, and 2-hour plasma glucose distributions in US population subgroups: NHANES 2005-2010. Ann Epidemiol. 2014; 24:83-89.*
12. *International Diabetes Federation. IDF Diabetes Atlas, 6th Ed. 2013. Available at https://www.idf.org/sites/default/files/EN_6E_Atlas_Full_0.pdf. Accessed 29 March 2015.*
13. *Gupta R, Gupta R, Deedwania PC, Sharma KK, Gupta A, Guptha S, Achari V, Asirvatham S, Bhansali A, Gupta BK, Gupta S, Jali MV, Mahanta TG, Maheshwari A, Saboo B, Singh J, Gupta R. Association of education, occupation and socioeconomic status with cardiovascular risk factors in Asian Indians: A cross-sectional study. PLoS One. 2012; 7:e044098.*
14. *Gupta A, Gupta R, Sharma KK, Achari V, Asirvatham AJ, Bhansali A, Gupta BK, Gupta S, Jali MV, Mahanta T, Maheshwari A, Saboo B, Singh J, Deedwania PC. Prevalence of diabetes and cardiovascular risk factors in middle-class urban populations in India. BMJ Open Diab Res Care. 2014; 2:e000048.*

15. *Luepkar RV, Evans A, McKeigue P, Reddy KS. Cardiovascular Survey Methods. 3rd Ed. Geneva. World Health Organization. 2002.*
16. *Gupta R, Sharma KK, Gupta BK, Gupta A, Saboo B, Maheshwari A, Mahanta TG, Deedwania PC. Geographic epidemiology of cardiometabolic risk factors in urban middle-class residents in India: A cross sectional study. J Global Health. 2015; 5:010411.*
17. *Guptha S, Gupta R, Deedwania P, Bhansali A, Maheshwari A, Gupta A, Gupta BK, Saboo B, Singh J, Achari V, Sharma KK. Cholesterol lipoproteins, triglycerides and prevalence of dyslipidemias among urban Asian Indian subjects: a cross sectional study. Indian Heart J. 2014; 66:280-288.*
18. *R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.*
19. *Martin Maechler (2015). diptest: Hartigan's Dip Test Statistic for Unimodality - Corrected. R package version 0.75-7. https://CRAN.R-project.org/package=diptest*

Table 1: Age-group specific prevalence and 95% confidence intervals of diabetes and impaired fasting glucose in males and females in the study cohort.

Females (n=2422)		Males (n=2937)		Numbers	Age groups
Impaired fasting glucose	Diabetes	Impaired fasting glucose	Diabetes	Males/Females	
9.8(7.4-12.2)	6.0(4.0-8.0)	13.7(11.2-16.2)	3.8(2.5-5.1)	417	**<30**
23.7(20.3-27.1)	8.9(6.5-11.3)	19.3 (16.4-22.2)	14.5 (12.1-16.9)	976	**30-39**
22.7(19.3-26.0)	18.6(15.4-21.8)	26.0(22.8-29.2)	27.6(24.6-30.6)	1445	**40-49**
27.4(23.8-31.0)	33.6(29.7-37.5)	25.1(21.9-28.3)	35.7(32.5-38.9)	1313	**50-59**
30.1(26.4-33.8)	36(32.0-40.0)	26.7(23.4-29.9)	41.0(37.7-44.3)	859	**60-69**
33.3(29.5-37.1)	42.1(38.0-46.2)	30.6(27.2-34.0)	40.8(37.5-44.1)	349	**70+**

Table 2: Mean, standard deviation, median and skewness of various anthropometric and biochemical variables in males (n=2937).

Kurtosis	Skewness	Interquartile intervals (25-75th percentile)	Median	Standard deviation	Mean	Variable
9.040	-0.891	9.000	167.000	8.006	166.314	**Height (cm)**
4.095	0.446	14.000	69.000	11.700	69.467	**Weight (kg)**
11.350	1.328	4.753	24.913	4.117	25.136	**Body mass index (kg/m2)**
3.926	0.156	14.000	90.000	10.693	89.931	**Waist (cm)**
6.061	0.098	0.072	0.960	0.067	0.956	**Waist-hip ratio**
4.471	0.398	0.080	0.538	0.065	0.541	**Waist-height ratio**
15.727	2.945	32.000	95.000	43.100	106.738	**Blood glucose (mg/dl)**
8.792	1.955	92.800	143.000	89.904	162.769	**Serum triglycerides (mg/dl)**

Table 3: Means, standard deviation, median and skewness of various anthropometric and biochemical variables in females (n=2422).

Kurtosis	Skewness	Interquartile intervals (25-75th percentile)	Median	Standard deviation	Mean	Variable
7.125	-0.836	9.000	154.000	8.694	154.024	**Height (cm)**
5.154	0.663	16.000	61.000	12.288	61.610	**Weight (kg)**
12.198	2.128	6.137	25.510	6.213	26.179	**Body mass index (kg/m2)**
3.125	0.029	14.000	86.000	11.239	85.172	**Waist (cm)**
4.537	0.001	0.106	0.924	0.081	0.914	**Waist-hip ratio**
3.235	0.251	0.101	0.553	0.078	0.555	**Waist-height ratio**
14.522	3.017	28.000	93.000	42.624	105.703	**Blood glucose (mg/dl)**
19.591	2.528	74.525	133.000	71.725	143.462	**Triglycerides (mg/dl)**

Table 4: Hartigans' dip test for unimodality

p-value	Value of D statistics	Variable
0	0.021813	**Fasting Glucose**
< 2.2e-16	0.025939	**Weight**
0.79496	0.004071	**Body mass index**
< 2.2e-16	0.025481	**Waist**
< 2.2e-16	0.027574	**Waist Hip Ratio**
0.00022	0.010269	**Waist Height Ratio**
0.395114	0.005091	**Triglycerides**
0.69149	0.004333	**Log triglycerides**
3.61E-06	0.012416	**HDL**
0.278212	0.005469	**Cholestrol**

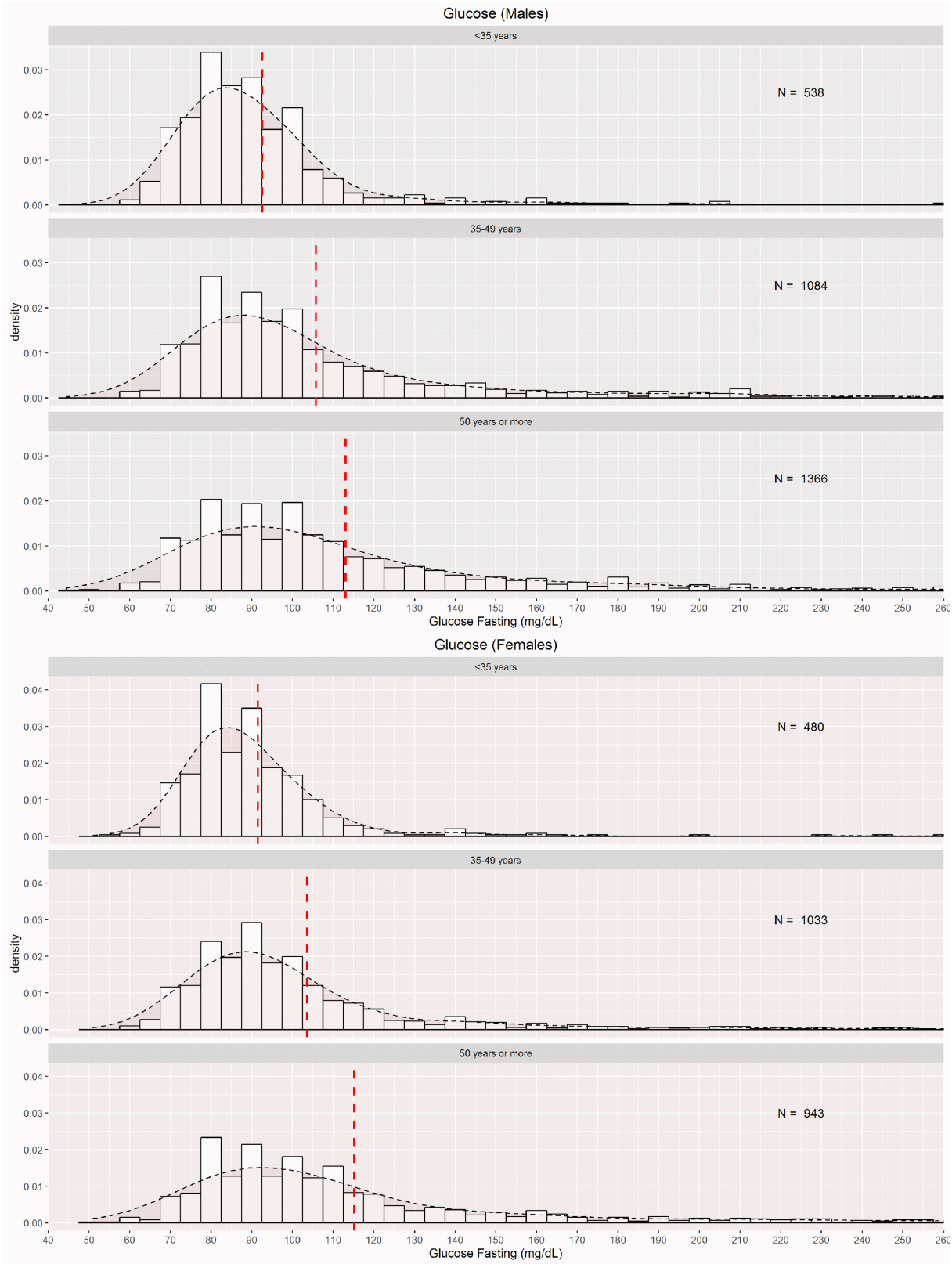

Figure 1: Fasting glucose distribution at various age-groups in males and females.

Figure 2: Waist size distribution at various age-groups in males and females.

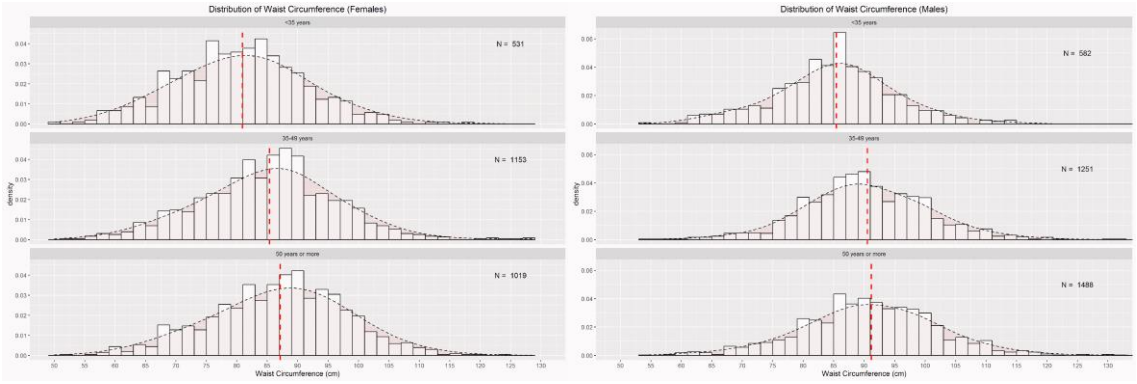

Figure 3: Waist-hip ratio distribution at various age-groups in males and females.

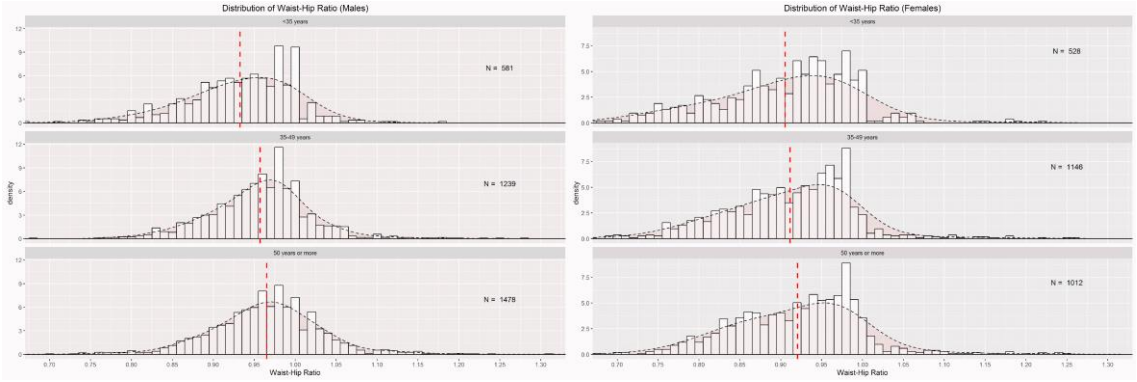

Figure 4: Blood triglycerides distribution at various age-groups in males and females.

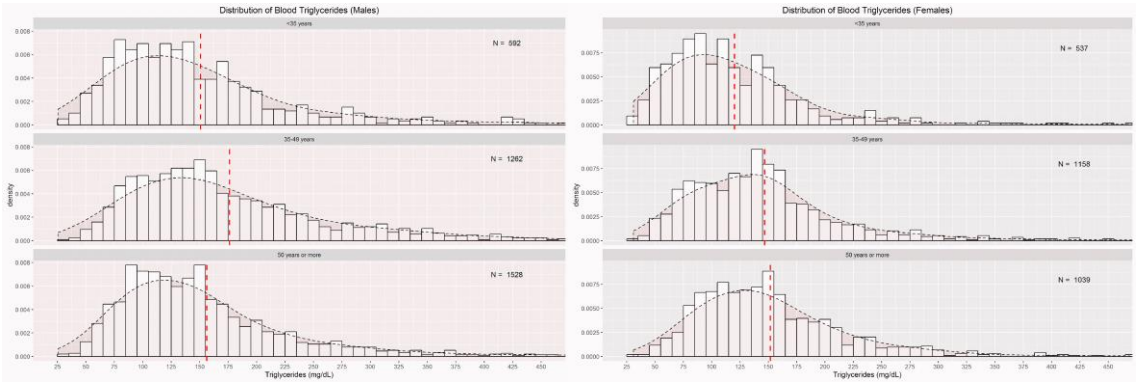

Figure 5: Body mass index distribution at various age-groups in males and females.