

Original Article

Evaluation of Thyroid Function Tests of Primary Hypothyroidism Diagnosed Cases' by Using Big Data: 13-Year Single Center ExperienceIbrahim Sahin¹, Canan Ersoy², Ilker Ercan³, Melahat Dirican^{4*}¹Department of Biostatistics, Institute of Health Sciences, Bursa Uludag University, Bursa, Turkey.²Department of Endocrinology and Metabolism, Faculty of Medicine, Bursa Uludag University, Bursa, Turkey.³Department of Biostatistics, Faculty of Medicine, Bursa Uludag University, Bursa, Turkey.⁴Department of Biochemistry, Faculty of Medicine, Bursa Uludag University, Bursa, Turkey.

ARTICLE INFO

ABSTRACT

Received 22.10.2020
Revised 23.11.2020
Accepted 01.12.2020
Published 10.12.2020

Key words:

big data;
hypothyroidism;
Triiodothyronine;
Thyroxine;
Thyrotropin

Objective: Our aim is to perform an analysis, using big data, of cases diagnosed with primary hypothyroidism and aged 18 and over who presented to our hospital, by evaluating the laboratory and socio-demographic data of the patients. Clustering analysis was performed in the big dataset for the purpose of structure-search study on the subject.

Methods: According to ICD-10 diagnoses of hypothyroidism between 2005-2018 in our hospital 130159 patients aged 18 and over with E03 and E06 diagnosis codes were included in the study. Since drugs containing levothyroxine used in primary hypothyroidism treatment have an effect on the measured hormone levels, in our study, TSH, fT3 and fT4 laboratory values in the first diagnosis of cases who had not received any treatment as part of the diagnosis according to demographics were analysed. Patients with one or more missing laboratory values were excluded, and data of 2680 patients with complete data and TSH values above 4.94 mU/L were retained. Analysis was made with the k-means clustering technique, with the data separated into two sets. k-means clustering was performed by including age, TSH, fT3 and fT4 variables. Cliff's Delta effect size coefficients and confidence intervals were calculated to perform size of the difference.

Results: The higher prevalence of primary hypothyroidism in female and the peak in hypothyroidism at 4-5 decades in both genders were observed. In which ages were low, fT3 and fT4 values were higher, whereas TSH values were lower in male. In which ages were low, TSH values were higher, whereas fT4 values were lower in female.

Conclusion: This study is the first big data analysis study carried out about primary hypothyroidism in our country. Despite the difficulties in implementation, it should not be forgotten that studies like these are important methods for enabling data to be created in our country.

Introduction

With the substantial increase seen in digitalisation and the use of technology, and the necessity to examine information density caused by data storage, there is a need for big

data studies in all areas. The concept of "Big Data", which has become important in recent years, is the process of making large amounts of data meaningful and workable.

In the field of healthcare, the first acceleration in big data applications appeared in the year

* Corresponding Author: Email: iesahin@outlook.com

2008. Although it was not often preferred prior to the year 2012, following that year, there was a significant increase in the number of published academic papers began to appear (1).

Since big data studies are, by their nature, concerned with the examination of data sets containing very large amounts of data, it is necessary for these data to be clustered to suit the purpose of the study.

Hypothyroidism is frequently observed as a result of dysfunction of the thyroid gland, and rarely, due to dysfunction of the pituitary gland or hypothalamus. Hypothyroidism, which occurs due to slowing of the functions of the thyroid gland itself, is named as primary hypothyroidism. The most common causes of primary hypothyroidism are iodine deficiency and Hashimoto's disease, which is an autoimmune thyroiditis. For diagnosis of primary hypothyroidism, the most sensitive method is to reveal higher thyroid stimulating hormone (TSH) and lower thyroid hormone levels in the blood (2,3) Examination of diseases related to thyroid dysfunction continues to increase in our country and globally day by day. All thyroid diseases are among the commonly occurring diseases, with a prevalence rate of 30% (4). The prevalence of overt hypothyroidism in the general population is reported to be between 0.2-5.3% in Europe and 0.3-3.7% in the USA (2) Despite having such a high prevalence, no comprehensive studies related to hypothyroidism have been conducted using big data in our country.

Our aim is to perform an analysis, using big data, of cases diagnosed with primary hypothyroidism and aged 18 and over who presented to our hospital, by evaluating the laboratory and socio-demographic characteristics of the patients.

Materials and Methods

The data used in the study were obtained by scanning the data processing SQL database at the Bursa Uludag University Health Application and Research Center. In order to carry out the study, permission was obtained from the Bursa Uludag University Ethics Committee to use patients' data related to their gender, age, date of admission to hospital, International Statistical Classification of Diseases and Related Health Problems and ICD-10 code, TSH, fT3 and fT4 laboratory values. (2018-05).

According to ICD-10 diagnoses of hypothyroidism patients accessible on the database for the years 2005-2018, the data of 130159 patients aged 18 and over with E03 and E06 diagnosis codes were included in the study.

Together with TSH measurement, which is the most sensitive method of primary hypothyroidism diagnosis, measurement of blood levels of free thyroxine (fT4) and, in necessary cases, free triiodothyronine (fT3) hormones is required (3). Since drugs containing levothyroxine used in primary hypothyroidism treatment have an effect on the measured hormone levels, in our study, TSH, fT3 and fT4 laboratory values in the first diagnosis of cases who had not received any treatment as part of the diagnosis were analysed. In addition to the laboratory data, demographic data related to age and gender variables were included in the data set. Patients with one or more missing laboratory measurement values were excluded, and data of 2,680 patients with complete data and TSH measurement values above 4.94 mU/L, which is the upper value of the normal range (Normal range=0.35-4.94) measured in the biochemistry laboratories at our hospital, were included in the study. The data were assessed

within the scope of big data as “male, female” (nm=539, nf=2141).

Statistical Analysis

Analysis was made with the k-means clustering technique, with the data separated into two sets. k-means clustering was performed by including age, TSH, fT3 and fT4 variables. Suitability of the data in the two clusters for normal distribution was tested with the Shapiro-Wilks test. Identification of the variables effective in the creation of the clusters was examined by performing effect size analyses. Effect size is defined as the size of the difference between groups in studies in which different interventions are compared (5). When there is an idea about the amount of variation existing within a group, effect size can be used as a measure for comparing this difference. Since the data was not normally distributed, Cliff's Delta effect size coefficients and confidence intervals were calculated as the effect size (6,7).

Cliff's Delta effect size coefficient should be interpreted respectively $|\delta| < 0.147$ as negligible effect size, $|\delta| < 0.330$ as small effect size, $|\delta| < 0.474$ as medium effect size and $|\delta| \geq 0.474$ as large effect size (8).

SQL was used at the stage of retrieval of the data from the database, Excel was used at the data organisation stage, and SPSS 22 software and the “effsize” package in R 3.5.3 software were used at the data implementation stage.

Results

20.1 % of the 2680 hypothyroid cases included in the study were male (n=539), while 79.9% of them were female (n=2141).

The histogram graph for the patients' age ranges according to gender is shown in Figure 1.

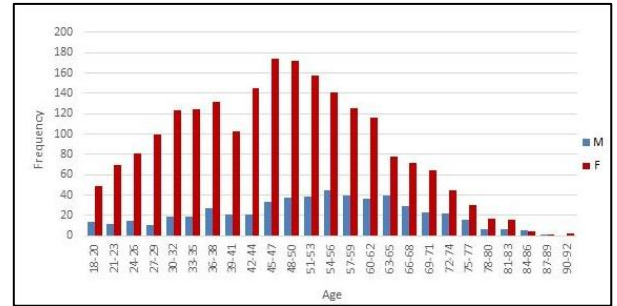


Figure 1. Histogram graph for the patients' age ranges according to gender

The higher prevalence of primary hypothyroidism in female and the peak in hypothyroidism at 4-5 decades in both genders. According to gender, cluster plots for age and TSH are shown in Figure 2 and Figure 3.

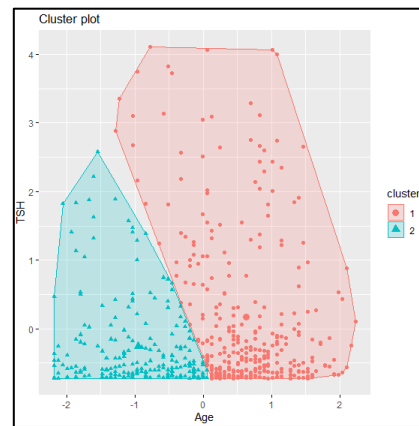


Figure 2. Cluster plot for male patients according to age and TSH

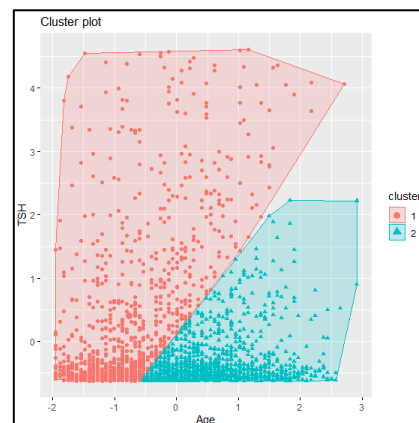


Figure 3. Cluster plot for female patients according to age and TSH

Evaluation of Data Related to Male Patients
Male patients were grouped in two clusters, and when the effect sizes were evaluated, a

large effect size was observed with regard to the age variable. However, it was seen that there was a medium effect size in terms of fT3. As for TSH and fT4, a small effect size

between the two clusters was observed. In the cluster in which ages were low, it was seen that fT3 and fT4 values were higher, whereas TSH values were lower (Table 1).

Table 1. Comparison of age and laboratory measurement values of male patients grouped into two clusters

	Cluster-1(n=214)	Cluster-2(n=325)	Cliff's Delta	
	Median (Min-Max.)	Median (Min-Max.)	δ	%95 C. I.
Age	37 (18-53)	61 (32-87)	0.944	0.915:0.963
TSH	8.62 (4.94-68.52)	12.41 (4.96-97.95)	0.217	0.120:0.310
fT3	2.78 (1.00-4.25)	2.25 (1.00-9.09)	0.407	0.313:0.495
fT4	0.93 (0.40-1.73)	0.81 (0.40-1.66)	0.250	0.153:0.342

Evaluation of Data Related to Female Patients
Female patients were grouped in two clusters, and when the effect sizes were evaluated, a large effect size for age, medium effect size for TSH, small effect size for fT4 was observed. However, it was seen that there was a negligible effect size in terms of fT3. In the cluster in which ages were low, it was seen that TSH values were higher, whereas fT4 values were lower (Table 2).

Discussion

Determining prevalence of diseases in society, making early diagnoses, and generating solutions related to treatment can be possible via examination of large numbers of data. For

this reason, there is a need for analysis using big data in the healthcare field, and studies conducted with regard to big data in the field of healthcare have gained importance (9-11). Besides the positive aspects of big data use, there are also aspects which may be negative. The data organisation process should be performed in a controlled way with the help of a subject matter expert and the data should be studied meticulously without being manipulated. Since a large number of data is being worked with, every incorrect operation that is made will result in very serious problems. The data set to be examined must be obtained from reliable sources.

Table 2. Comparison of age and laboratory measurement values of female patients grouped into two clusters

	Cluster-1(n=968)	Cluster-2(n=1173)	Cliff's Delta	
	Median (Min-Max.)	Median (Min-Max.)	δ	%95 C. I.
Age	34 (18-87)	55 (39-90)	0.826	0.798:0.851
TSH	11.69 (4.95-99.12)	7.46 (4.94-56.24)	0.351	0.306:0.396
fT3	2.53 (1.00-4.96)	2.55 (1.00-8.43)	0.012	-0.038:0.062
fT4	0.88 (0.40-2.44)	0.95 (0.40-2.34)	0.213	0.164:0.261

Clustering analysis was performed in the big data set for the purpose of structure-search study on the subject.

In our study, it was seen that 79.9% of cases diagnosed with primary hypothyroidism within the scope of big data analysis were women.

In our study, it was seen that the number of male diagnosed with primary hypothyroidism showed an increase up to the age of 54, and that the highest number of cases was reached at the ages of 54-56. The number of cases was observed to decrease after the age of 56.

It was also seen that the number of female diagnosed with primary hypothyroidism showed an increase up to the age of 45, and that the highest number of cases was reached at the ages of 45-47. The number of cases was observed to decrease after the age of 47.

When the male cases in our study were examined, it was seen that fT3 and fT4 values were higher, whereas TSH values were lower. When we made an assessment according to data structure in which there was differentiation in male cases, it was seen that in cases in the 18-53 age range (median 37), TSH ranged between 4.94-68.52 (median 8.62), fT3 values ranged between 1.00-4.25 (median 2.78) and fT4 values ranged between 0.40-1.73 (median 0.93). In cases in the 32-87 age range (median 61), TSH ranged between 4.96-97.95 (median 12.41), fT3 values ranged between 1.00-9.09 (median 2.25) and fT4 values ranged between 0.40-1.66 (median 0.81).

When the female cases in our study were examined, it was seen that TSH values were higher, whereas fT4 values were lower. There was a negligible effect size in terms of fT3. When we made an assessment according to data structure in which there was differentiation in female cases, it was seen that in cases in the 18-87 age range (median 34),

TSH ranged between 4.95-99.12 (median 11.69) and fT4 values ranged between 0.40-2.44 (median 0.88). In cases in the 39-90 age range (median 55), TSH ranged between 4.94-56.24 (median 7.46) and fT4 values ranged between 0.40-2.34 (median 0.95).

When thyroid diseases, which are commonly seen in the world and in our country, are examined with regard to gender, it is seen that they are more common in women (2,12). In a number of studies, it is determined that the rate of thyroid diseases between female and male is between 3.5 and 12. Female predominance in hyperthyroidism, hypothyroidism, overt and subclinical hypothyroidism diseases is observed in regions where low and high iodine consumption occurs (2,13). The female/male ratio in our study was calculated as 3.97, and a significant percentage of patients was seen to consist of women.

Similarly, thyroid disease could be more commonly seen with increasing age (14). The highest frequencies were reached at ages 48-50 in women and 54-56 in men. 27.98% of women were aged between 45-55, and similarly, 25.79% of men were aged between 45-55.

In the study conducted by Takeda et al. (15) with 1.007 Japanese patients, it was concluded that in male, there was a negative relationship between age and values for the fT4 measurement variable. TSH, however, was determined to have a positive relationship with age in women and men separately. In our study, however, while similar results were obtained for TSH and fT4 values in male, higher TSH values were observed in female cases at lower ages.

When Jammah et al. (16) examined the data as a whole, a negative correlation between fT4 and age was determined, while no statistical relationship was found between TSH and age.

In our study, however, in men, similar values were observed for fT4 according to age, while lower values were observed for TSH in cases at lower ages. In women, on the other hand, in cases at lower ages, higher values were observed for TSH, whereas lower values were seen for fT4.

When the results of our study are examined, the higher prevalence of primary hypothyroidism in female and the peak in hypothyroidism at 4-5 decades in both genders are seen to correspond with the literature. In our study, as expected, TSH level increased in male as age increased, whereas, contrary to the literature, it decreased in female. Iodine deficiency is a common cause of hypothyroidism in our country. In women in the period of fertility, due to increased use of iodine by the foetus and excretions of iodine in urine during pregnancy, if sufficient iodine is not supplemented, iodine deficiency deepens and TSH value can increase. This situation can explain why TSH was higher in individuals at younger ages in the fertility period in women included in our study.

In cases of primary hypothyroidism, insufficient T3 and T4 hormone is synthesised from the thyroid gland. As an attempt is made to preserve the level of T3, which is the more active of the hormones, in the body, T4 is the first to decrease. The fewer the thyroid hormones secreted by the thyroid gland are, the greater the TSH stimulation from the pituitary gland will be. In our study, T3 and T4 values were found to be lower in groups with high TSH in both genders, as expected.

Conclusion

Big data studies are very important for evaluating data for commonly seen diseases in our country in terms of revealing age, gender, prevalence, factors affecting development of

the disease, diagnosis and treatment outcomes. To enable these results to be used in clinical practice and to increase data reliability, the correct entering of ICD-10 diagnosis codes, accurate input of diagnosis data and correct recording of early diagnosis, and complete records of laboratory data and treatment practices are important and necessary.

In conclusion, this study is the first big data analysis study carried out on the subject of primary hypothyroidism in our country. Despite the difficulties in implementation, it should not be forgotten that studies like these are an important method for enabling data to be created in our country.

Ethics Committee Approval

In order to carry out the study, permission was obtained from the Bursa Uludag University Ethics Committee to use patients' data related to their gender, age, date of admission to hospital, International Statistical Classification of Diseases and Related Health Problems and ICD-10 code, TSH, fT3 and fT4 laboratory values. (2018-05).

Conflict of Interest

No conflict of interest was declared by the authors.

Financial Disclosure

The authors declared that this study received no financial support.

References

1. Baro E, Degoul S, Beuscart R, Chazard E. Toward a Literature-Driven Definition of Big Data in Healthcare. *Biomed Res Int.* 2015;2015:639021. <https://doi.org/10.1155/2015/639021>
2. Taylor PN, Albrecht D, Scholz A, Gutierrez-Buey G, Lazarus JH, Dayan CM, et al. Global epidemiology of hyperthyroidism and hypothyroidism. *Nat Rev Endocrinol.*

- 2018;14(5):p 301-316.
<https://doi.org/10.1038/nrendo.2018.18>
3. The Society of Endocrinology and Metabolism of Turkey, Thyroid Diseases Diagnosis and Treatment Guideline. 2019,4th ed., Ankara,Turkish Clinics,40-50.
 4. Serin OS, İlhan M, Ahcı S, Okuturlar Y, Koc G, Eyupgiller T ve ark. The level of awareness on thyroid disorders. The Medical Bulletin of Şişli Etfal Hospital, 2006; 50(3): 181-185.
<https://dx.doi.org/10.5350/SEMB.20160412042738>
 5. Cohen J. Statistical Power Analysis for the Behavioral Sciences 1988; 8-14, 2nd ed. Hillsdale, NJ: Erlbaum.
 6. Cliff N. Dominance statistics: Ordinal analyses to answer ordinal questions. Psychological Bulletin 1993;114(3): 494–509.
<https://doi.org/10.1037/0033-2909.114.3.494>
 7. Hogarty KY, Kromrey JD. Using SAS to Calculate Tests of Cliff's Delta. Proceedings of the Twenty-Fourth Annual SAS User Group International Conference, Miami, Florida, 1999;238.
 8. Romano J, Kromrey JD, Coraggio J, Skowronek J. Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys? Annual Meeting of the Florida Association of Institutional Research, 2006;1-33.
 9. Grasso MA, Comer AC, DiRenzo DD, Yesha Y, Rishe ND. Using Big Data to Evaluate the Association between Periodontal Disease and Rheumatoid Arthritis. AMIA Annu Symp Proc,2015; 589-593.
<https://www.ncbi.nlm.nih.gov/pubmed/26958193>
 10. Jeong K, Lee JD, Kang DR, Lee S. A population-based epidemiological study of anaphylaxis using national big data in Korea: trends in age-specific prevalence and epinephrine use in 2010–2014. Allergy, Asthma & Clinical Immunology, 2018; 14(31):1-9. <https://doi.org/10.1186/s13223-018-0251-z>
 11. Kumari S. Breast Cancer Classification Using Big Data Approach. PARIPEX - Indian Journal of Research, 2018; 7(1): 401-403.
 12. Hueston W, Carek P, Allweiss, P. Endocrine Disorders in Chapter 35. Current Diagnosis & Treatment Family Medicine 2nd Edition, 2008; 392, McGraw-Hill Companies.
 13. Iglesias P, Diez J. Hypothyroidism in Male Patients: A Descriptive, Observational and Cross-Sectional Study in a Series of 260 Men. The American Journal of the Medical Sciences, 2008; 336(4): 315-320.
<https://doi.org/10.1097/MAJ.0b013e318167b0d0>
 14. Levy EG. Thyroid Disease in the Elderly. Medical Clinics of North America, 1991; 75(1): 151-167.
[https://doi.org/10.1016/s0025-7125\(16\)30476-x](https://doi.org/10.1016/s0025-7125(16)30476-x)
 15. Takeda K, Mishiba M, Sugiura H, Nakajima A, Kohama M, Hiramatsu S. Evaluated Reference Intervals for Serum Free Thyroxine and Thyrotropin Using the Conventional Outliner Rejection Test without Regard to Presence of Thyroid Antibodies and Prevalence of Thyroid Dysfunction in Japanese Subjects. Endocrine Journal, 2009; 56(9): 1059-1066.
<https://doi.org/10.1507/endocrj.K09E-123>
 16. Jammah AA, Alshehri AS, Alrakhis AA, Alhedaihy AS, Almadhi AM, Alkwai HM, et al. Characterization of thyroid function and antithyroid antibody tests among Saudis. Saudi Medical Journal, 2015; 36(6): 692-7.
<https://www.ncbi.nlm.nih.gov/pubmed/25987111>