Research Paper

Presentation Bootstrap Resampling Method in ROC Curve Analysis to Assessment Diagnostic Tests of TSH Serum in Neonatal Hypothyroidism

Ali Akbarpour Sohbatabadi[1], Azam Moslemi[2], Fatemeh Dorreh[3], Mohammad Rafiei[4*]

1. MSc of Biostatistics, Department of Biostatistics, School of Medical Sciences, Arak University of Medical Sciences, Arak, Iran
2. Assistant of Biostatistics, Department of Biostatistics, School of Medical Sciences, Arak University of Medical Sciences, Arak, Iran
3. Associate professor of Pediatric, Department of Pediatric, School of Medical Sciences, Arak University of Medical Sciences, Arak, Iran
4. Professor of Biostatistics, Department of Biostatistics, School of Medical Sciences, Arak University of Medical Sciences, Arak, Iran

Received: 2020/01/28 **Accepted:** 2020/04/22

Abstract

Background: Congenital Hypothyroidism is one of the most prevalence disease of endocrinology and metabolic infant and important caused of mental deficiency. Hypothyroidism in the secondary causes represented by problem in pituitary or hypothalamus and shortage of TSH. Clinical symptom of Hypothyroidism at the first days of birth are indiscernible so only way of diagnosis and treatment of that, is screening. In this study we used the bootstrap method to evaluation the ROC curve of diagnostic screening TSH test.

Materials and Methods: In this study a bootstrap method is proposed to plot ROC curve and determinate the cut point of TSH & T4 based on YOUDEN method. at the first generated subsample and estimated ROC and cut-off point for each sample. The procedure run for 500 replicate. For each replicate selected cut-off where have maximum YOUDEN. The cut-off point calculated for 500 replications from each sample and then selected mean of 500 replicate. The cut-off which have minimum misclassification selected optimal cut-off point.

Results: Based ordinary method to estimate the ROC curve the optimal cut-off point for TSH variable is 10 mlU/L and for T4 variable is 100 µg/dL which based the bootstrap method respectively optimal cut-off point for TSH is 9.54 and for T4 is 100.07.

Conclusion: The optimal cut-off point where obtained from bootstrap method have minimum misclassification rate versus ordinary method. Based the small sample bootstrap have minimum misclassification rate toward ordinary method with large sample.

***Corresponding Author:** Mohammad Rafiei
Address: Department of Biostatistics, School of Medicine, Arak University of Medical Sciences, Sardasht, Basij Square, Arak, Markazi Province, Iran
Tel: +98-8634173520
E-mail: rafeie@arakmu.ac.ir

Keywords: Bootstrap method, Hypothyroidism, ROC curve, Sensitivity, Specificity.

How to cite this article: Akbarpour Sohbatabadi A., Moslemi A., Dorreh F., Rafiei M. Presentation Bootstrap Resampling Method in ROC Curve Analysis to Assessment Diagnostic Tests of TSH Serum in Neonatal Hypothyroidism, Bootstrap Resampling to Assessment Diagnostic Tests of TSH, Journal of Sabzevar University of Medical Sciences, 2022; 29(1):51-64.

Copyright © 2022 Sabzevar University of Medical Sciences. This work is licensed under a Creative Commons Attribution- Non Commercial 4.0 International license (https://creativecommons.org/licenses/by-nc/4.0/). Non-commercial uses of the work are permitted, provided the original work is properly cited.
Published by Sabzevar University of Medical Sciences.

Introduction

Congenital hypothyroidism is one of the most common endocrine and metabolic diseases in infants and one of the important and preventable causes of mental retardation in infants. In Iran, the incidence of hypothyroidism is estimated at 2.2 per thousand live births, which is higher than other countries in the world. The incidence of this disease in female infants is twice that of male infants. Delay or lack of treatment leads to severe growth disorders and irreversible mental retardation. Loss of IQ due to this disease can be prevented when the disease is diagnosed in a short time (ideally in the first days of life). According to the national screening, the incidence of the disease in Iran is about 2.4% which is higher than the global incidence. Screening goals include treating infected infants in the first days of life, diagnosing, treating transient types of disease, and creating a suitable setting for screening for other neonatal metabolic diseases. The incidence of this disease in female infants is twice that of male infants. Delay in treatment or lack of treatment leads to severe developmental disorders and irreversible mental retardation. Loss of IQ due to this disease can be prevented when the disease is diagnosed in a short time and ideally in the first days of life. In Iran, according to the national guidelines for screening for congenital hypothyroidism, Newborns are referred to their health centers for screening in 3-5 days after birth, Thyrotropin (TSH) is measured using a heel blood sample collected on filter paper. Due to the strategy for diagnostic tests, the prevalence of the disease has increased in recent years and mild cases of C.H. are also diagnosed. It seems in addition to the geographical area, the cut-off point for TSH is also related to the diagnosis of hypothyroidism and the incidence of the disease. Children with TSH higher than normal are called to test for infected children by measuring serum thyroxine T4 and TSH hormones and to treat and follow up. Use of ROC curve is an effective and well-known method for evaluating a diagnostic test whose results are ordinal or quantitative variables. Bootstrap resampling is a method that regardless of many hypotheses, by creating numerous samples, approaches sample conditions to population conditions. With this method, by considering all types of sample formation, it is possible to estimate the population by coefficients and confidence intervals with the least bias and error. In using this method, we will not need to accept the assumptions related to the normality of the distribution and the existence of a large sample. Bootstrap is one of the most powerful estimation tools and the accuracy of the estimates under Bootstrap is relatively accurate. But not many studies have been done on the ROC curve and bootstrap method. In this study, the bootstrap method for diagnostic tests of TSH and T4 is used and simulated for real samples.

Methodology

Sample consists of 614 infants who referred to health centers for screening at the age of 3-5 days. They were selected and examined from 2006 to the end of 2012 in Arak city and screening centers in rural areas. The infants included in the study were from Amir Kabir, Quds, Imam Khomeini and Taleghani hospitals and special rural screening centers. Sampling of infants' heels were performed on 903S&S filter paper and sent to the provincial health center along with no.1 screening form. TSH levels were measured by ELISA antibody kits. Infants with TSH more than 5 were recalled and venous samples were taken from them for T4, T3RU and TSH measurement. According to the national guidelines, infants with TSH more than 10 mIU/L and T4 less than 6.5 µg/dL were considered as patients. Finally, the sample formed with 614 members, including 414 unhealthy and 200 healthy infants. then ROC curve cut-off point was drawn and the optimal cut-off points for TSH and T4 variables were obtained based on this curve. Using cut-off point and a statistical model, sick infants can be distinguished from healthy ones. The function of cut-off point is based on probability, in this way that by using a statistical model e.g. logistic model, the probability of being sick is calculated for all infants. Any of infants with a higher probability than the cut-off point is considered as patient and any of them with a lower probability than cut-off point is considered healthy. Chi-square and independent t-test were used to analyze the data and ROC curve tests and diagnostic test based on Youden index were used to test the research hypotheses. The Youden Index was introduced in 1950 by Yuden. This index was proposed as a way to summarize the performance of a diagnostic test and its value varies from 0 to 1, and the closer this index to 1, is higher the efficiency and test detection. The bootstrap resampling method was performed with 500 replications and different sample sizes (n=10, 25, 75, 100) based on the following algorithm with umber of replicates (N)

a) In every replication, the following steps are performed:

1. Sample selection with replacement (size: n)
2. Estimating the sensitivity and specificity of the sample

3. Drawing ROC curve based on logistic model

b) Determining the "optimal cut-off point" by calculating the average of N cut-off points
c) Classification of patients based on the optimal cut-off point
d) Calculating the accuracy of the test based on the optimal cut-off point

In this study, the number of replicates was N=500 and the sample sizes were n=200, 150, 100, 75, 25, 10. Using the Bootstrap sampling method, the mean disease proportion and confidence interval were calculated and compared with the disease proportion of real data. Finally, the exact value of the optimal cut-off points for the TSH and T4 were obtained by calculating the average of 500 optimal points obtained from 500 replications per sample size. Data analysis was performed by SAS 9.4 software.

Result

Based ordinary method to estimate the ROC curve the optimal cut-off point for TSH variable is 10 mlU/L and for T4 variable is 100 µg/dL. Results showed that the area of under Roc curve (AUC) for the TSH per sample size indicated the desirability of the diagnostic test of hypothyroidism prediction. Table 4 showed that the length of the confidence interval and the mean standard deviations reduced as the increasing sample size. It indicated the accuracy of the diagnostic test at large sample sizes.

4. Selection of the cut-off point based on the maximum value of the Youden index

Based on the results obtained from the bootstrap method, it can be concluded that the TSH test is appropriate. The bootstrap method confirmed the results obtained from the ROC curve, which reached most accuracy in the large sub-samples. The results in the sample size of 25 were almost similar to the results obtained from the ROC curve with the sample size of 612, which indicates that the bootstrap method is suitable for small sample sizes.

Based on Table 4, the estimated cut-off points of the bootstrap method for sample size of 200, 150 and 100 were equal. So, the cut-off point was not depended on the sample size and this point was the best cut-off point for neonatal hypothyroidism. The lowest classification error rate was 0.022 for n=100 and the cut-off point of 9.55. As a result, the cut-off point and the classification error rate was constant as increasing the sample size.

In addition, based on the bootstrap method for the T4, classification error rate was 0.18 for sample sizes less than 75, equal 100 and 150. Table 6 indicated that the cut-off point was 100.44, 100.07 and 101 for sample sizes 75, 100 and 150, respectively. So, lowest the classification error rate observed in sample size 100. The best cut-off point was 100.07.

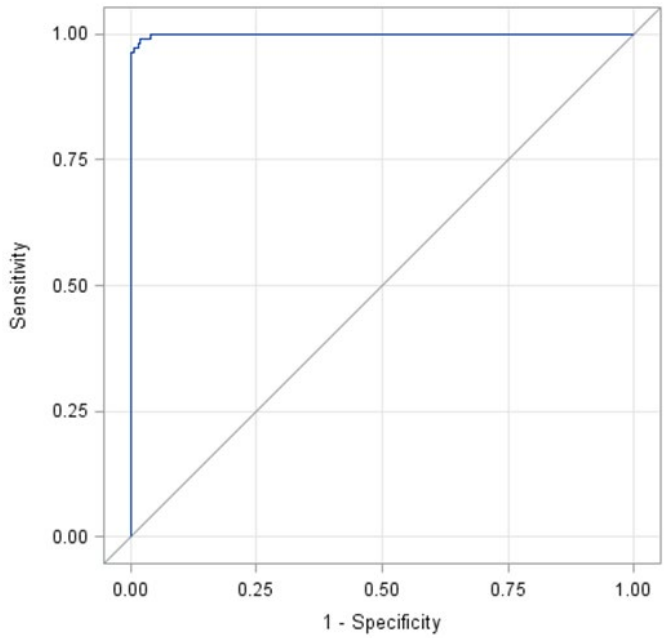


Figure 1. ROC curve for TSH

Table 4. Result of Bootstrap method for TSH with 500 replicates

Sample size	Sensitivity	specificity	AUC	Std. deviation	Confidence interval for AUC (95%)		Optimal cut off point	misclassification rate
					Lower band	Upper band		
10	0.94	0.96	0.9992	0.0035	0.9904	1.0043	9.48	0.04
25	0.94	0.97	0.999	0.0024	0.9927	1.002	9.43	0.03
75	0.95	0.98	0.9991	0.001	0.9953	0.9994	9.5	0.03
100	0.97	0.98	0.9991	0.0008	0.9957	0.9989	9.55	0.02
150	0.96	0.98	0.9991	0.0007	0.996	0.9987	9.55	0.02
200	0.96	0.98	0.9991	0.0005	0.9963	0.9985	9.55	0.02

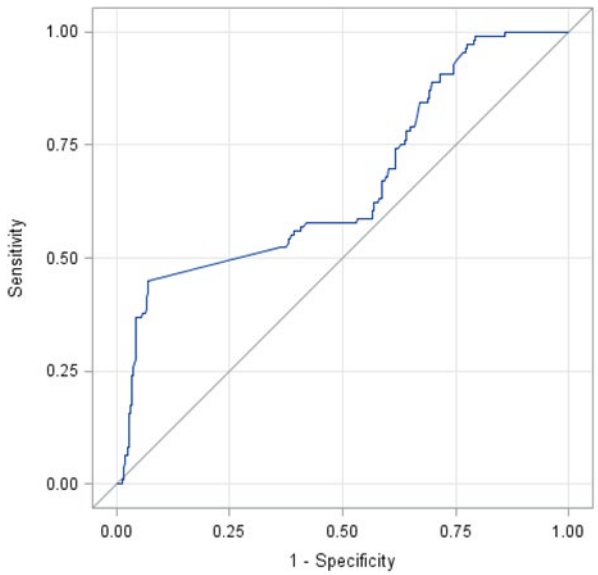


Figure 2. ROC curve for T4

Table 6. Result of Bootstrap method for T4 with 500 replicates

Sample size	Sensitivity	specificity	AUC	Std. deviation	Confidence interval for AUC (95%)		Optimal cut off point	misclassification rate
					Lower band	Upper band		
10	0.53	0.54	0.6536	0.1560	0.4190	0.9140	78.07	0.42
25	0.54	0.55	0.6660	0.0980	0.4750	0.8670	90.57	0.41
75	0.56	0.57	0.6750	0.0484	0.5712	0.7615	100.44	0.18
100	0.55	0.58	0.6850	0.0381	0.5850	0.7480	100.07	0.18
150	0.54	0.56	0.7010	0.0080	0.6010	0.7200	101.00	0.18
200	0.54	0.56	0.7120	0.0070	0.6130	0.7260	99.30	0.39

Conclusions

In the present study, based on the ROC curve, the cut-off point for TSH was 10 mlU/L, and at this point the sensitivity, specificity and AUC were 98.83%, 100% and 98.99%, respectively. The cut-off point was obtained 9.55 mlU/L based on the bootstrap method in a sample size of 100. For T4, using the ROC curve and the bootstrap method, the cut-off point was 100 µg/dl. There was not much difference in the amount of classification error rate between the ROC curve and the bootstrap method, which confirms the cut-off point obtained by the bootstrap method. The optimal cut-off point where obtained from

bootstrap method have minimum misclassification rate versus ordinary method. Based the small sample bootstrap have minimum misclassification rate toward ordinary method with large sample.
Studies show that when the required sample size is available, the estimation error is significantly reduced by increasing the bootstrap iterations as well as increasing the bootstrap sample size. According to the results, this is also true for the amount of sensitivity and specificity. Due to the nature of the bootstrap method for estimation, if sufficient sample size is not available (n< 25), bootstrap is a suitable solution to establish the conditions that can be used to obtain a reliable estimate and in low sample size this method can Be a good solution as we have seen in different sample sizes, the results of the bootstrap method indicate a valid cut-off point for the ROC curve and can be used for validation.

Acknowledgment

The present article was extracted from an MSc thesis and was financially supported by Arak University of Medical Sciences (IR.ARAKMU.REC.1398.140).

Conflict of Interest

The authors declare that there are no conflict of interest regarding the publication of this manuscript.

مقاله پژوهشی

ارائه روش بازنمونه‌گیری بوت‌استرپ در تحلیل منحنی ROC به‌منظور ارزیابی آزمون تشخیصی TSH و T4 سرمی در کم‌کاری تیروئید نوزادان

علی اکبرپور صحبت‌آبادی[۱]، اعظم مسلمی[۲]، فاطمه درّه‌ای[۳] محمد رفیعی[۴*]

۱. کارشناس‌ارشد آمار زیستی، دانشکده پزشکی، دانشگاه علوم پزشکی اراک، اراک، ایران
۲. استادیار آمار زیستی، گروه آمار زیستی، دانشکده پزشکی، دانشگاه علوم پزشکی اراک، اراک، ایران
۳. دانشیار گروه اطفال، گروه اطفال، دانشکده پزشکی، دانشگاه علوم پزشکی اراک، اراک، ایران
۴. استاد آمار زیستی، گروه آمار زیستی، دانشکده پزشکی، دانشگاه علوم پزشکی اراک، اراک، ایران

تاریخ دریافت: ۱۳۹۸/۱۱/۰۸ تاریخ پذیرش: ۱۳۹۹/۰۲/۰۳

* نویسندهٔ مسئول: محمد رفیعی
نشانی: مرکزی، اراک، سردشت، میدان بسیج، دانشگاه علوم پزشکی اراک، دانشکده پزشکی
تلفن: ۰۹۱۸۸۶۳۷۴۴۵
رایانامه: rafeie@arakmu.ac.ir
شناسه ORCID: 0000-0002-7373-510x
شناسه ORCID نویسنده اول: 0000-0003-3965-8945

چکیده

زمینه و هدف: کم‌کاری مادرزادی تیروئید از شایع‌ترین بیماری‌های غدد درون‌ریز و متابولیک کودکان و از مهم‌ترین علل عقب‌ماندگی ذهنی نوزادان است. کم‌کاری این غده می‌تواند در علل ثانویه به شکل نقص کارکرد در هیپوفیز یا هیپوتالاموس و کمبود TSH باشد. علائم بالینی کم‌کاری تیروئید در روزهای اول زندگی کم و غیرقابل تشخیص است؛ بنابراین تنها راه تشخیص و درمان مؤثر انجام غربالگری می‌باشد. در این مطالعه با استفاده از منحنی ROC و روش بوت استرپ، قدرت تشخیص آزمایش غربالگری TSH و T4 ارزیابی شد.

مواد و روش‌ها: در این مطالعه، روش بوت استرپ برای ترسیم منحنی ROC و تعیین نقطه برش TSH و T4 براساس شاخص Youden پیشنهاد شد. ابتدا زیرنمونه‌های مختلف، تولید و برای هر نمونه نقطه برش و نمودار ROC محاسبه شد. روش بالا ۵۰۰ بار تکرار شد و در هر تکرار نقطه برشی انتخاب شد که دارای ماکزیمم شاخص Youden بود. برای هر نمونه میانگین ۵۰۰ تکرار نقطه برش محاسبه شد و نقطه برشی با مینیمم خطای دسته‌بندی را داشت به‌عنوان نقطه برش بهینه انتخاب شد.

یافته‌ها: براساس منحنی ROC نقطه برش بهینه برای متغیر TSH برابر ۱۰ میلی واحد در لیتر و برای متغیر T4 برابر ۱۰۰ میکروگرم در دسی‌لیتر بود. با استفاده از روش بوت استرپ مقدار دقیق نقطه برش برای متغیر TSH، مقدار ۹.۵۴ m/l و برای متغیر T4، مقدار ۱۰۱.۵ mg/dl به‌دست آمده بود.

نتیجه‌گیری: نقطه برش بهینه‌ای که از روش بوت استرپ به‌دست آمد خطای دسته‌بندی کمتری نسبت به منحنی ROC دارد. در حجم نمونه‌های کوچک نیز خطای دسته‌بندی پایینی نسبت به منحنی ROC نسبت به حجم نمونه بزرگ مشاهده شد.

کلیدواژه‌ها:
حساسیت، ویژگی، منحنی ROC، بوت استرپ، کم‌کاری تیروئید

۱. مقدمه

کم‌کاری مادرزادی تیروئید[1] از شایع‌ترین بیماری‌های غدد درون‌ریز و متابولیک نوزادان و از علل مهم و قابل پیشگیری عقب‌ماندگی ذهنی در نوزادان محسوب می‌شود. میزان بروز کم‌کاری تیروئید در کشورهای توسعه‌یافته از جمله فرانسه یک در هزار تولد زنده، چین ۰/۴۹ در هزار تولد زنده و یونان ۱/۲۵ در هزار تولد زنده برآورد گردیده است (۱). در ایران میزان بروز کم کاری تیروئید ۲/۲ در هزار تولد زنده برآورد شده است که نسبت به سایر کشورهای جهان، بروز بالاتری دارد (۲). میزان بروز این بیماری در نوزادان دختر دو برابر نوزادان پسر بیان شده

1 Congenital Hypothyroidism

Copyright © 2022 Sabzevar University of Medical Sciences. This work is licensed under a Creative Commons Attribution- Non Commercial 4.0 International license (https://creativecommons.org/licenses/by-nc/4.0/). Non-commercial uses of the work are permitted, provided the original work is properly cited.
Published by Sabzevar University of Medical Sciences.

است (۳). تأخیر در درمان یا درمان نشدن به اختلالات شدید در رشد و عقب‌ماندگی ذهنی غیرقابل برگشت منجر می‌شود. از دست دادن ضریب هوشی ناشی از این بیماری زمانی قابل پیشگیری است که بیماری در مدت‌زمان کوتاه و به‌طور ایده‌آل در اولین روزهای زندگی تشخیص داده شود (۴, ۵). براساس غربالگری کشوری، بروز بیماری در ایران حدود ۲/۴ درصد و بالاتر از آمار جهانی است. اهداف غربالگری شامل درمان نوزادان مبتلا در روزهای اول تولد، تشخیص، درمان انواع گذرای بیماری و ایجاد بستری مناسب برای غربالگری دیگر بیماری‌های متابولیک نوزادان است (۶, ۷).

در ایران براساس دستورالعمل کشوری، غربالگری کم‌کاری مادرزادی تیروئید، نوزادان در ۵-۳ روز پس از تولد برای غربالگری به مراکز بهداشتی درمانی خود مراجعه می‌کنند و نمونه خون پاشنه پا روی کاغذ فیلتر جمع‌آوری می‌گردد و هورمون تیروتروپین[1] (TSH) اندازه گیری می‌شود (۴, ۸). با توجه به استراتژی برای آزمون‌های بیماریابی در سال‌های اخیر، شیوع بیماری افزایش یافته است و موارد خفیف C.H[2] نیز تشخیص داده می‌شود. به‌نظر می‌رسد علاوه بر منطقه جغرافیایی، نقطه برش تعیین شده برای TSH نیز در شناسایی و میزان بروز بیماری دخیل باشد. کودکان با TSH بالاتر از حد نرمال، فراخوان می شوند تا با سنجش هورمون‌های تیروکسین T4و TSH سرم، کودکان مبتلا مشخص شود و تحت درمان و پیگیری قرار گیرند (۴, ۸).

یک روش مؤثر و شناخته شده برای ارزیابی یک آزمون تشخیصی که نتایج آن متغیری در مقیاس رتبه‌ای یا کمی باشد، استفاده از منحنی ROC[3] می‌باشد (۹, ۱۰).

روش بازنمونه‌گیری بوت استرپ، روشی است که فارغ از بسیاری از فرضیات، با ایجاد نمونه‌های فراوان، شرایط نمونه را به شرایط جامعه نزدیک می‌کند و با در نظر گرفتن تمامی حالات تشکیل نمونه می‌توان برآورد ضرایب و برآورد فاصله اطمینان را با کم‌ترین میزان اریبی و خطا برای برآورد جامعه در نظر گرفت (۱۱).

در استفاده از این روش، نیازمند پذیرش پیش‌فرض‌های مربوط به نرمال بودن توزیع و وجود نمونه بزرگ نخواهیم بود (۱۲). بوت استرپ یکی از قوی‌ترین ابزارهای برآورد می‌باشد و صحت برآوردها تحت بوت استرپ به‌طور نسبتاً دقیقی بیان می‌شود. اما مطالعات زیادی در زمینه منحنی ROC و بوت استرپ انجام نشده است؛ بنابراین در این مطالعه، روش بوت استرپ برای آزمون تشخیصی TSH و T4 شرح داده و برای نمونه‌های واقعی شبیه‌سازی می‌شود.

۲. مواد و روش‌ها

نمونه‌ای متشکل از ۶۱۴ نوزاد که در سن ۵-۳ روزگی به مراکز بهداشت برای غربالگری مراجعه کردند، انتخاب شدند و از سال ۸۵ تا پایان سال ۹۱ در شهرستان اراک و مراکز غربالگری در روستاها آزمایش شدند. نوزادان از بیمارستان‌های امیرکبیر، قدس، امام خمینی، طالقانی و مراکز مخصوص غربالگری روستایی وارد مطالعه شدند. نمونه‌گیری از پاشنه پای نوزادان بر کاغذ فیلتر با نام S&S-۹۰۳ انجام و همراه فرم شماره یک غربالگری به مرکز بهداشت استان ارسال شد. میزان TSH به‌وسیله کیت‌های پادتن علم به روش ELIZA سنجیده شد. نوزادان با TSH بیشتر از پنج فراخوان شدند و از آنها نمونه وریدی به‌منظور سنجش T4, TSH, T3RU گرفته شد. با توجه به دستورالعمل کشوری نوزاد با TSH بیش از ۱۰ میلی واحد در لیتر و T4 کمتر از ۶/۵ میکروگرم در دسی‌لیتر بیمار در نظر گرفته شد. در نهایت نمونه‌ای متشکل از ۶۱۴ نوزاد شامل ۴۱۴ تن آنها بیمار و ۲۰۰ تن آنها سالم تشکیل شد و نقطه برش منحنی ROC برای نوزادان، رسم و نقطه برش بهینه برای متغیر TSH و T4 براساس این منحنی به‌دست آمد.

با استفاده از نقطه برش و یک مدل آماری می‌توان افراد بیمار را از سالم متمایز کرد. عملکرد نقطه برش براساس احتمال، بدین صورت است که با استفاده از مدل آماری مانند مدل لجستیک، احتمال بیمار بودن همه افراد محاسبه می‌شود. هر فردی با احتمال بیمار بودن بیشتر از نقطه برش، بیمار در نظر گرفته می‌شود و هر فردی با احتمال بیمار بودن کمتر از نقطه برش، سالم در نظر گرفته می‌شود.

برای تجزیه‌وتحلیل داده‌ها از آزمون کای اسکوئر و تی مستقل استفاده شد و برای بررسی فرضیات پژوهش از آزمون‌های منحنیROC و آزمون تشخیصی بر پایه شاخص Youden استفاده شد. شاخص Youden را یودن در سال ۱۹۵۰ معرفی کرد. این شاخص به‌عنوان روشی برای خلاصه کردن عملکرد یک آزمون تشخیصی پیشنهاد شد و مقدار آن از ۰ تا ۱ متغیر است و هرچه این شاخص به سمت ۱ نزدیک‌تر باشد میزان کارایی و تشخیص آزمون بالاتر است. مقدار آن از طریق فرمول زیر محاسبه می‌شود: (۱۲)

1 Thyroid Stimulating Hormone
2 Congenital Hypothyroidism
3 Receiver operating characteristic

j= sensitivity + specificity -1

روش بازنمونه‌گیری بوت استرپ با تعداد تکرار ۵۰۰ و حجم نمونه‌های مختلف (10، 25، 75، n=100) براساس الگوریتم زیر اجرا شد.

تعداد N تکرار داریم.

الف- در هر تکرار مراحل زیر انجام می‌شود:

۱- انتخاب نمونه با جای‌گذاری به حجم n

۲- برآورد میزان حساسیت و ویژگی آن نمونه

۳- رسم منحنی ROC براساس مدل لجستیک

۴- انتخاب نقطه برش براساس ماکزیمم شاخص Youden

ب- تعیین نقطه برش بهینه از طریق محاسبه میانگین N نقاط برش

ج- دسته‌بندی بیماران براساس نقطه برش بهینه

د- محاسبه دقت تشخیص تست براساس نقطه برش بهینه

در این پژوهش تعداد تکرار N=500 و حجم نمونه 10، 25، 75، 100، ۱۵۰، ۲۰۰ =n در نظر گرفته شد. با استفاده از روش بازنمونه‌گیری بوت استرپ، میانگین نسبت موفقیت‌ها و فاصله اطمینان محاسبه و با نسبت موفقیت داده‌های واقعی مقایسه گردید. در نهایت، مقدار دقیق نقطه برش بهینه برای متغیر TSH و T4 با محاسبه میانگین ۵۰۰ نقطه بهینه به‌دست‌آمده از ۵۰۰ تکرار به‌ازای هر حجم نمونه به‌دست آمد. تجزیه‌وتحلیل داده‌ها با استفاده از نرم‌افزار SAS 9.4 انجام شد.

۳. یافته‌ها

ابتدا به بررسی مشخصات دموگرافیک پرداخته شد، سپس حساسیت، ویژگی و منحنی ROC برای متغیر TSH رسم گردید و در حالت بوت استرپ برای تکرارهای مختلف انجام شد.

جدول ۱. بررسی توزیع جنسیت در گروه سالم و بیمار

وضعیت بیماری / جنسیت	تعداد بیمار	درصد	تعداد سالم	درصد	کل
دختر	۲۰۶	٪۶۷	۱۰۰	۳۳	۳۰۶ (٪۱۰۰)
پسر	۲۰۸	٪۶۷	۱۰۰	۳۳	۳۰۸ (٪۱۰۰)
جمع	۴۱۴		۲۰۰		
	$\chi^2 = 0.003$		سطح معنی داری =۰.۹۵		

با توجه به جدول ۱ و آزمون خی دو ملاحظه شد بین جنسیت و ابتلا به بیماری کم‌کاری تیروئید ارتباط معنی‌داری وجود ندارد (P>0.05)

جدول ۲. بررسی مقایسه میانگین بین میزان TSH و T4 در دو گروه سالم و بیمار

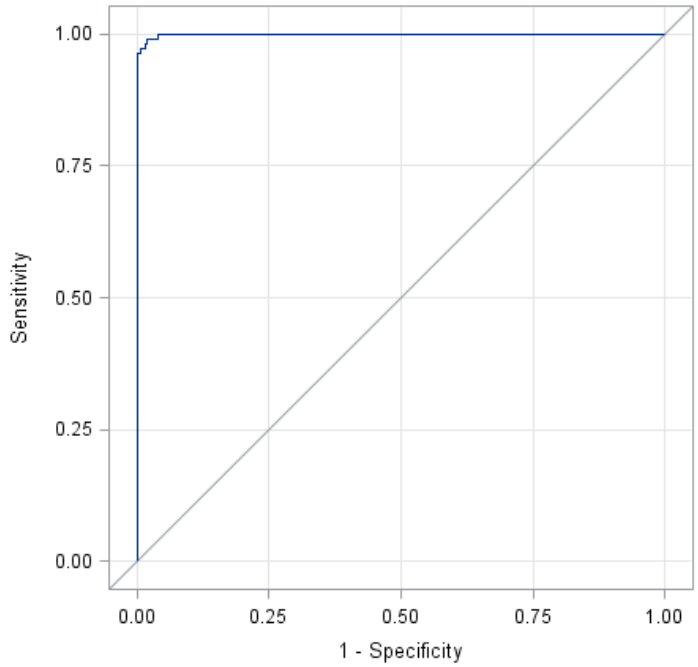
متغیر	سالم	بیمار	t	p
TSH	۲.۱۵±۴	۲۸.۲۹±۳۵	۱۵.۳۷	p<۰.۰۰۱
T4	۵۸.۱۲±۷۵.۸۸	۴۷.۱۲±۵۴.۲۷	-۳.۲۵	p<۰.۰۰۱

با توجه به مقدار سطح معناداری در جدول ۲ اختلاف معنی‌داری بین مقدار TSH و T4 در دو گروه سالم و بیمار مشاهده شد (p<0.05)

جدول ۳. حساسیت و ویژگی برای متغیر TSH بر اساس منحنی ROC

متغیر	حساسیت	ویژگی	انحراف استاندارد	فاصله اطمینان		نقطه برش	نقطه برش بر اساس احتمال	سطح زیر منحنی (AUC)
TSH	٪۹۵	٪۹۷.۸	۰.۲۷۸۳	-۲.۱۵۸۹	-۱.۰۳۲۸	۱۰	۰.۷۳۷۹	۰.۹۹۷۴

براســاس جدول ۳ ســطح زیر منحنی ROC برای متغیر مقدار TSH، میزان ۰/۹۹ به‌دســت آمد که نشــان‌دهنده قوی بودن آزمون تشــخیصــی TSH می‌باشــد. خطای دســته‌بندی براساس منحنی ROC برابر ۰/۰۲۶ به‌دست آمد.



نمودار ۱. منحنی ROC برای متغیر TSH

نتایج روش بوت استرپ برای متغیر TSH بعد از ۵۰۰ تکرار در جدول ۴ نشان داده شده است.

جدول ۴. به‌کار بردن روش بوت استرپ برای متغیر TSH

حجم زیر نمونه	حساسیت	ویژگی	میانگین سطح زیر منحنی	میانگین انحراف استاندارد	فاصله اطمینان AUC		نقطه برش بهینه	خطای دسته‌بندی
					حد پایین	حد بالا		
۱۰	٪۹۴	٪۹۶	۰.۹۹۹۱۹۷	۰.۰۰۳۵۲۱	۰.۹۹۰۴	۱.۰۰۴۳	۹.۴۷۹	۰.۰۳۵
۲۵	٪۹۳.۵	٪۹۷	۰.۹۹۹۰۳۱	۰.۰۰۲۳۶۶	۰.۹۹۲۷	۱.۰۰۲۰	۹.۴۳	۰.۰۲۷
۷۵	٪۹۵	٪۹۸	۰.۹۹۹۰۵۹	۰.۰۰۱۰۳۱	۰.۹۹۵۳	۰.۹۹۹۴	۹.۵۰	۰.۰۲۹
۱۰۰	٪۹۶.۵	٪۹۷.۸	۰.۹۹۹۰۸۲	۰.۰۰۰۸۱۴	۰.۹۹۵۷	۰.۹۹۸۹	۹.۵۵	۰.۰۲۲
۱۵۰	٪۹۶	٪۹۸.۳	۰.۹۹۹۰۹	۰.۰۰۰۶۶	۰.۹۹۶۰	۰.۹۹۸۷	۹.۵۵	۰.۰۲۲
۲۰۰	٪۹۶	٪۹۸	۰.۹۹۹۱	۰.۰۰۰۵۱	۰.۹۹۶۳	۰.۹۹۸۵	۹.۵۵	۰.۰۲۲

در جدول ۴ ملاحظه می‌شود مقدار به‌دست‌آمده سطح زیر منحنی متغیر TSH بــه‌ازای حجم نمونــه‌هــای مختلف، نشــان‌دهنده مطلوب بودن آزمون تشــخیصــی برای پیش‌بینی بیماری کم‌کاری تیروئید اســت. با افزایش حجم نمونه، طول فاصله اطمینان کوتاه‌تر و میانگین انحرافات استاندارد کوچک‌تر می‌شود که نشان‌دهنده دقیق بودن آزمون تشخیصی در حجم نمونه بالا اســت. با توجه به نتایج به‌دســت‌آمده از روش بوت اســترپ، می‌توان نتیجه گرفت که آزمون تشــخیص TSH، مناسب است. براساس نتایج به‌دست‌آمده، روش بوت استرپ تأییدی بر نتایج به‌دســت‌آمده از منحنی ROC اســت که این نتیجه در انتخاب زیرنمونه‌های بالاتر به دقیق‌ترین میزان خود خواهد رسید. در جدول ۴ ملاحظه می‌شود که نتایج در حجم نمونه ۲۵ تایی تقریباً مشــابه نتایج به‌دســت‌آمده از منحنی

ROC با نمونه ۶۱۲ تایی می‌باشد که نشان‌دهنده این است روش بوت استرپ، در حجم نمونه کم، مناسب است.

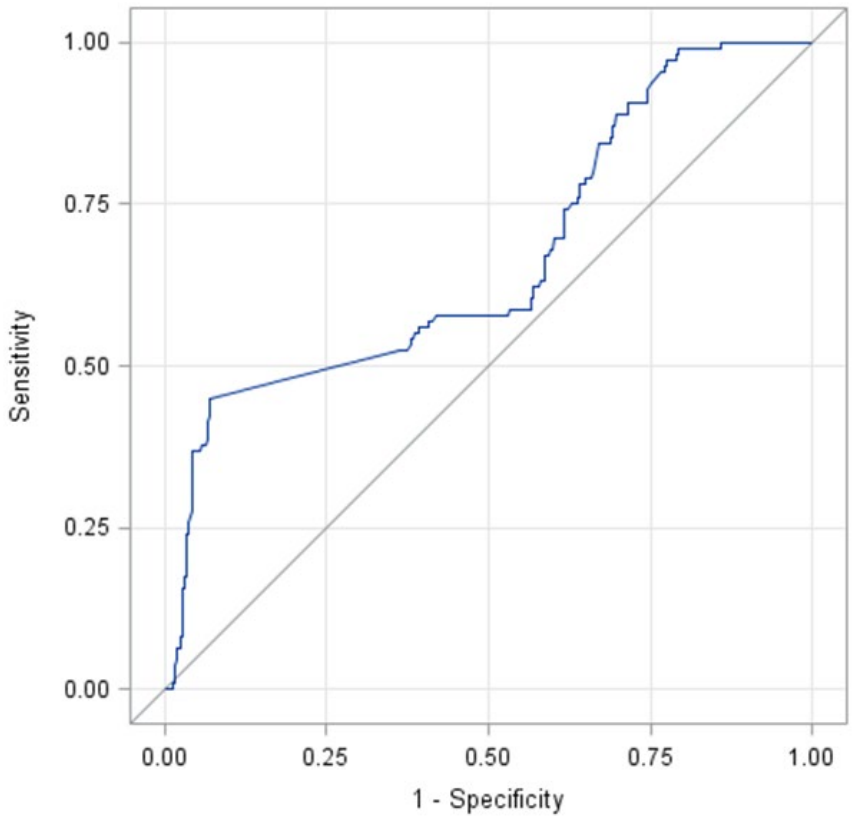
در برآورد نقطه برش‌ها ملاحظه می‌شود که در حجم نمونه‌های ۱۰۰، ۱۵۰ و ۲۰۰ تایی نقطه برش‌های یکسانی به‌دست آمده است که نشان‌دهنده این موضوع است که نقطه برش به‌دست‌آمده دیگر به حجم نمونه بستگی نداشته و این نقطه، بهترین نقطه برش با استفاده از روش بوت استرپ برای کم‌کاری تیروئید نوزادان است.

در زمانی که n=100 برابر ۰/۰۲۲ است مقدار خطای دسته‌بندی به کمترین مقدار خود رسیده و نقطه برش ۹.۵۵ کمترین میزان خطای دسته‌بندی را داشته است. در نتیجه طبق جدول ۲ نیز مشاهده می‌شود که با افزایش نمونه‌ها نقطه برش ثابت می‌باشد و خطای دسته‌بندی نیز ثابت است.

جدول ۵. حساسیت و ویژگی برای متغیر T4 براساس منحنی ROC

متغیر	حساسیت	ویژگی	انحراف استاندارد	فاصله اطمینان		نقطه برش	نقطه برش بر اساس احتمال	سطح زیر منحنی (AUC)
T4	٪۵۶.۹	٪۵۷.۸	۰.۰۰۲۱	۰.۰۰۳۸	۰.۰۱۲۴	۱۰۰	۰.۶۹۷	۰.۶۶۶۳

براساس جدول ۵ سطح زیر منحنی ROC برای متغیر مقدار T4، میزان ۰/۶۶ به‌دست آمد که نشان‌دهنده ضعیف بودن آزمون تشخیصی T4 می‌باشد. خطای دسته‌بندی براساس منحنی ROC برابر ۰/۱۸ به‌دست آمد.



نمودار ۲. منحنی ROC برای متغیر T4

نتایج روش بوت استرپ برای متغیر T4 بعد از ۵۰۰ تکرار در جدول ۶ نشان داده شده است.

جدول ۶. به‌کار بردن روش بوت استرپ برای متغیر T4

تعداد نمونه	حساسیت	ویژگی	میانگین سطح زیر منحنی	میانگین انحراف استاندارد	فاصله اطمینان AUC		نقطه برش	خطای دسته‌بندی
					حد پایین	حد بالا		
۱۰	٪۵۳	٪۵۴.۲	۰.۶۵۳۶	۰.۱۵۶	۰.۴۱۹	۰.۹۱۴	۷۸.۰۷	۰.۴۲
۲۵	٪۵۴	٪۵۵	۰.۶۶۶	۰.۰۹۸	۰.۴۷۵	۰.۸۶۷	۹۰.۵۷	۰.۴۱
۷۵	٪۵۶	٪۵۷	۰.۶۷۵	۰.۰۴۸۴۱۱	۰.۵۷۱۲۲۶	۰.۷۶۱۴۵۴	۱۰۰.۴۴	۰.۱۸
۱۰۰	٪۵۵.۲	٪۵۷.۸	۰.۶۸۵	۰.۰۳۸۱	۰.۵۸۵	۰.۷۴۸	۱۰۰.۰۷	۰.۱۸
۱۵۰	٪۵۴.۱	٪۵۶	۰.۷۰۱	۰.۰۰۸	۰.۶۰۱	۰.۷۲۰	۱۰۱	۰.۱۸
۲۰۰	٪۵۴	٪۵۶	۰.۷۱۲	۰.۰۰۷	۰.۶۱۳	۰.۷۲۶	۹۹.۳	۰.۳۹

با توجه به جدول ۶ و استفاده از روش بوت استرپ برای متغیر T4 نقطه برش بعد از نمونه‌های ۷۵ تایی خطای دسته‌بندی ۰/۱۸ داشت همچنین در نمونه‌های ۱۰۰ و ۱۵۰ تایی نیز خطای دسته‌بندی ۰/۱۸ به‌دست آمد. در این نمونه‌ها ابتدا در نمونه ۷۵ تایی نقطه برش ۱۰۰/۴۴ و در نمونه ۱۰۰ تایی نقطه برش برابر ۱۰۰/۰۷ و در نمونه ۱۵۰ تایی نیز نقطه برش برابر ۱۰۱ به‌دست آمد که خطای دسته‌بندی در همسایگی نمونه ۱۰۰ مناسب‌ترین نقطه برش را دارد؛ زیرا هرچه حجم نمونه بالا و پایین نقطه شود؛ نقطه برش نیز تغییر می‌کند که همان‌طور مشاهده شد در نمونه ۲۰۰ تایی خطای ۰/۳۹ و در نمونه ۲۵ تایی خطای ۰/۴۱ مشاهده شد؛ بنابراین بهترین نقطه برش برابر۱۰۰/۰۷ به‌دست آمد. در ارتباط با متغیر T4، ۶۵ درصد داده‌ها مقادیر گمشده بود. انحراف معیار و پراکندگی این متغیر نیز بالا بود که می‌تواند موجب یافتن نقطه برش نامناسب باشد.

٤. بحث و نتیجه‌گیری

در مطالعه حاضر براساس منحنی ROC نقطه برش برابر ۱۰ میلی واحد در لیتر به‌دست آمد و در این نقطه، مقدار حساسیت و ویژگی و مساحت زیر نمودار به‌ترتیب برابر ۹۸/۸۳ درصد، ۱۰۰ درصد و ۹۸/۹۹ درصد گزارش شد.

نقطه برش، براساس روش بازنمونه‌گیری بوت استرپ در حجم نمونه ۱۰۰، ۹.۵۵ میلی واحد در لیتر به‌دست آمد. در مطالعه سیلوستری[1] و همکاران (۲۰۱۷) نقطه برش برای TSH سرمی نوزادان ۱۰/۲ میلی واحد در لیتر و مقدار حساسیت ۱۰۰ درصد و ویژگی ۹۳ درصد به‌دست آمد که با مطالعه حاضر هم‌خوانی داشت (۱۳). در مطالعه تی یو[2] و همکاران نیز نقطه برش ۱۰ میلی واحد در لیتر برای سطح TSH سرمی در نوزادان و مقدار حساسیت و ویژگی در این نقطه به‌ترتیب برابر ۹۸/۹۳ درصد و ۹۹/۴۸ درصد به‌دست آمد که با نتایج مطالعه حاضر مطابقت دارد (۱۴). در مطالعه‌ای که در ترکیه انجام شد نقطه برش برای ۲۰میلی واحد در لیتر فراخوان شدند (۱۵). در مطالعه توصیفی دره و همکارن که با عنوان «ارتباط میزان فراخوان با وقوع کم‌کاری مادرزادی تیروئید در برنامه غربالگری کم‌کاری تیروئید نوزادان در شهرستان اراک» در سال ۱۳۸۵ انجام شد میزان نقطه برش ۶/۵ میلی واحد در لیتر به‌دست آمد (۱۶) که با مطالعه حاضر، اختلاف داشت و با توجه به مقدار نقطه برش‌های متفاوت گزارش شده در مطالعات قبلی برای متغیر TSH، در مطالعه حاضر روش بوت استرپ، بالاترین میزان حساسیت و ویژگی را در بین مطالعات گفته شده داشت و نقطه برشی که از این روش به‌دست آمد بسیار کاراتر و قوی‌تر است و می‌توان به‌عنوان نقطه برش جدیدی برای متغیر TSH در نظر گرفته شود.

در مطالعه حاضر با توجه به مقدار سطح زیر منحنی ROC برای متغیر TSH و روش بوت استرپ، TSH می‌تواند تنها و بهترین آزمون تشخیصی برای این بیماری باشد که نتایج این مطالعه با مطالعه شهان[3] و همکاران مطابقت دارد (۱۷).

همان‌طور که در جدول ۴ مشاهده شد مقدار خطای دسته‌بندی در حجم نمونه‌های بالا مقدار ثابت می‌باشد و به کمترین میزان خود رسیده است و نقطه‌های برش در این نمونه‌ها نیز تفاوت چندانی نداشتند که این نشان‌دهنده بهینه بودن نقطه برش به‌دست‌آمده برای متغیر TSH است.

1 Stela Maris Silvestrin
2 Tu WJ
3 Sheehan

برای متغیر T4 با استفاده از منحنی ROC نقطه برش نقطه برش برابر ۱۰۰ میکروگرم در دسی‌لیتر و با استفاده از روش بوت استرپ نیز تقریباً همان نقطه برش به‌دست آمد و در مقدار خطای دسته‌بندی اختلاف زیادی بین روش نمودار ROC و روش بوت استرپ دیده نشد که این امر نشان‌دهنده تأیید نقطه برش به‌دست‌آمده از طریق روش بوت استرپ است که در مطالعه دلبرت[1] و همکاران این نقطه برش ۷۵ اعلام شد که این نقطه برش با نتیجه مطالعه حاضر اختلاف جزیی در مقدار خطای دسته‌بندی داشت (۱۸). مقدار حساسیت و ویژگی در مطالعه حاضر برای متغیر T4 به‌ترتیب برابر ۵۷ درصد و ۵۸/۶ درصد گزارش شد که نسبت به موارد مشابه کار شده، حساسیت و ویژگی ضعیفی داشت. وجود درصد بزرگی داده گمشده و انحراف معیار بالای این متغیر می‌تواند دلیلی برای یافتن نقطه برش نامناسب باشد.

بررسی‌ها نشان می‌دهد زمانی که حجم نمونه موردنیاز در دسترس است با افزایش تکرارهای بوت استرپ و همچنین افزایش حجم نمونه بوت استرپ، خطای برآورد به میزان قابل‌توجهی کاهش می‌یابد. با توجه به نتایج به‌دست‌آمده این امر در مورد مقدار حساسیت و ویژگی نیز صادق است.

با توجه به ماهیت روش بوت استرپ برای برآورد، در صورتی که حجم نمونه کافی در دسترس نباشد (n<۲۵) بوت استرپ راه‌حل مناسبی برای برقرار کردن شرایطی است که بتوان برآورد معتبری را به‌دست آورد و در حجم نمونه کم این روش می‌تواند راه‌حل مناسبی باشد. همان‌طور که در حجم نمونه‌های مختلف دیدیم نتایج روش بوت استرپ نشان‌دهنده نقطه برش معتبر برای منحنی ROC است و می‌تواند برای معتبرسازی استفاده شود.

شاخص میانگین نسبت به مقادیر پرت و دور افتاده، حساس است اما شاخص میانه این حساسیت را ندارد. در مطالعه حاضر نقاط برش براساس میانگین برای TSH و T4 به‌دست آمد؛ بنابراین پیشنهاد می‌شود برای بهبود نقاط برش، مطالعه‌ای با استفاده از شاخص میانه تکرارهای روش بوت استرپ در نمودار ROC انجام گیرد و با نتیجه مطالعه حاضر مقایسه شود. همچنین مطالعه‌ای در زمینه نمودار ROC با استفاده از روش اعتبارسنجی[2] انجام شود و با روش بوت استرپ مقایسه گردد.

تقدیر و تشکر

بدین‌وسیله از معاونت محترم پژوهشی دانشگاه علوم پزشکی اراک و کلیه کسانی که در انجام این پژوهش ما را یاری کردند، تشکر و قدردانی می‌گردد. کد اخلاق این پژوهش IR.ARAKMU.REC.1398.140 است.

References

[1]. Deng K, He C, Zhu J, Liang J, Li X, Xie X, et al. Incidence of congenital hypothyroidism in China: data from the national newborn screening program, 2013–2015. Journal of Pediatric Endocrinology and Metabolism. 2018;31(6):601-8.
[2]. Osooli M, Haghdoost A, Yarahmadi S, Foruzanfar M, Dini M, Holakouie Naieni K. Spatial distribution of congenital hypothyroidism in Iran using geographic information system. Iranian Journal of Epidemiology. 2009;5(1):1-8.
[3]. BF III SG, Behrman R. Nelson text book of pediatrics. 19th edt. Philadelphia: Elsevier Saunders Company; 2011.
[4]. Ordookhani A, Mirmiran P, Hedayati M, Hajipour R, Azizi F. An interim report of the pilot study of screening for congenital hypothyroidism in Tehran and Damavand using cord blood spot samples. European journal of pediatrics. 2003;162(3):202-3.
[5]. Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian pediatrics. 2011;48(4):277-87.
[6]. Olivieri A, Hypothyroidism ISGfC. Epidemiology of congenital hypothyroidism: what can be deduced from the Italian registry of infants with congenital hypothyroidism. The Journal of Maternal-Fetal & Neonatal Medicine. 2012;25(sup5):7-9.
[7]. Eftekhari N, Gh A, Khaksari M, Salari Z. The prevalence rate of congenital hypothyroidism in Kerman/Iran in 2005-2007. Journal of Kerman University of Medical Sciences. 2015.
[8]. Dorreh F, Yousefi CP, Javaheri J, Eshrati B, Amiri Z. Evaluation of 6 years performance of screening program of congenital hypothyroidism in Markazi province (2006-2012)." (2013): 0-0.
[9]. Gallas BD. One-shot estimate of MRMC variance: AUC. Academic radiology. 2006;13(3):353-62.
[10].Narkhede S. Understanding auc-roc curve. Towards Data Science. 2018 Jun;26:220-7.
[11].Jokinen J, Nordström A-L, Nordström P. ROC analysis of dexamethasone suppression test threshold in suicide prediction after attempted suicide. Journal of affective disorders. 2008;106(1-2):145-52.
[12].Martínez-Camblor P, Pardo-Fernández JC. The Youden index in the generalized receiver operating characteristic curve context. The international journal of biostatistics. 2019 May 1;15(1).
[13].Silvestrin SM, Leone C, Leone CR. Detecting congenital hypothyroidism with newborn screening: the relevance of thyroid-stimulating hormone cutoff values. Jornal de pediatria. 2017;93(3):274-80.
[14].PKU C. Newborn screening for inborn errors of metabolism in Beijing, China: 22 years of experience. Journal of medical screening. 2011;18(4).

1 Delbert

2 Cross validation

[15].Simsek E, Karabay M, Kocabay K. Neonatal screening for congenital hypothyroidism in West Black Sea area, Turkey. International journal of clinical practice. 2005;59(3):336-41.
[16].Doreh F, MOHAMMADI T. The relationship between recall rate and the incidence of congenital hypothyroidism in the screening program for neonatal hypothyroidism in Arak, 2006. 2010.
[17].Sheehan MT. Biochemical Testing of the Thyroid: TSH is the Best and, Oftentimes, Only Test Needed - A Review for Primary Care. Clin Med Res. 2016;14(2):83–92. doi:10.3121/cmr.2016.1309.
[18].Delbert A, Fisher M. Disorders of the thyroid in the newborn and infant,(chapter 6) in Sperling MA, editor. Pediatric Endocrinology. Saunders. Elsevier; 2008.