

Juvenilinės hipertenzijos prognozavimo modeliai

Vytautas JANILIONIS (KTU), Edita JANAVIČIŪTĖ (KTU),
Jūratė KLUMBIENĖ (KMU), Giedrė SAKALAUSKIENĖ (KMU)
el. paštas: vytautas.janilionis@ktu.lt

1. Įvadas

Padidėjusio kraujospūdžio liga – arterinė hipertenzija (AH) išlieka svarbi daugelio šalių gyventojų sveikatos problema, nes ji labai paplitusi ir yra viena pagrindinių širdies ir kraujagyslių sistemos ligų rizikos veiksnių [1]. Lietuvoje širdies ir kraujagyslių sistemos ligos sudaro daugiau nei pusę visų gyventojų mirties priežasčių [2].

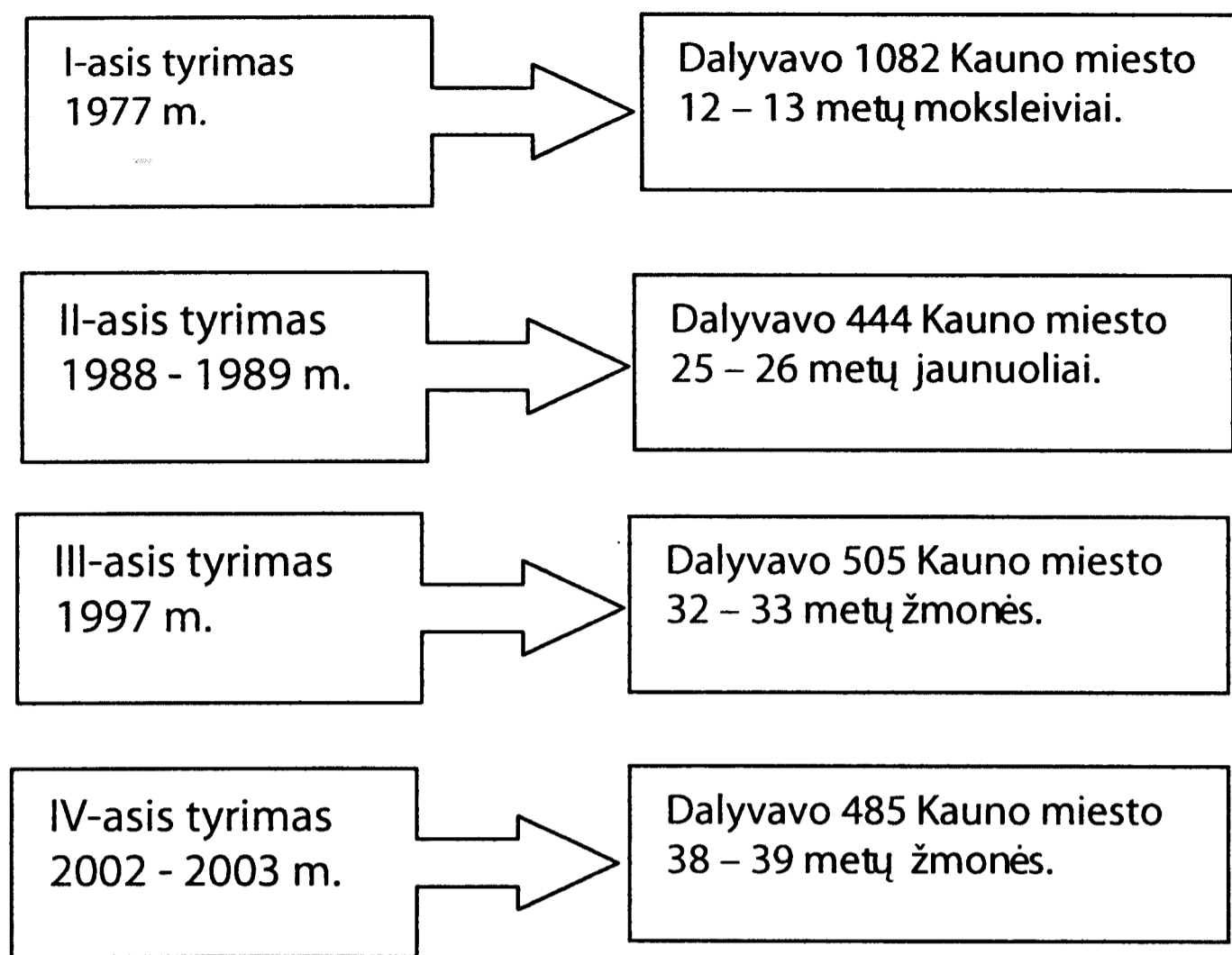
Pasaulyje atlikta daug suaugusiųjų AH epidemiologinių tyrimų. Literatūros analizė parodė, kad jų duomenų analizei dažniausiai taikomi koreliacinės analizės, daugialypės regresinės, logistinės regresinės analizės bei diskriminantinės analizės metodai [2–4]. Vaikų ir paauglių AH tirta gerokai rečiau. Dauguma juvenilinės (jaunatvinės) arterinės hipertenzijos tyrimų buvo trumpalaikiai, juose dalyvavo nedaug žmonių, todėl, pagal gautus rezultatus, negalima pateikti apibendrinančių išvadų [1]. Turimos žinios apie arterinio kraujospūdžio (AKS) raidą parodė, kad net kūdikystėje ar vaikystėje nustatyti kai kurie sveikatos rodikliai gali būti svarbūs AH atsiradimui vyresniame amžiuje [2]. Labai svarbus uždavinys – veiksnių, lemiančių suaugusiųjų AH atsiradimą, išaiškinimas ir ankstyva diagnostika, t.y., svarbu jau vaikystėje nustatyti rodiklius, kurie lemia AH atsiradimą vyresniame amžiuje ir prognozuoti AH atsiradimą.

KMU (Kauno medicinos universiteto) Biomedicininių tyrimų institute (BMTI) 1977 metais buvo pradėtas juvenilinės arterinės hipertenzijos tyrimas. Šios mokslinės programos dalyviai pakartotinai buvo tirti 3 kartus (1989, 1997 ir 2002/03 metais) (1 pav.). Sveikatos patikrinimų metu visiems vaikams (12–13 metų), jaunuoliams (25–26 metų) ir suaugusiems (32–33 bei 38–39 metų) buvo užpildoma standartinė apklausos anketa ir registruojami įvairūs parametrai: AKS, širdies susitraukimų dažnis, ūgis, svoris, užrašoma elektrokardiograma, širdies ultragarsinio bei biocheminio tyrimo rezultatai ir t.t.

Pagrindinis šio darbo tikslas – panaudojus atliktų tyrimų duomenis ir statistikos metodus sudaryti juvenilinės hipertenzijos atsiradimo prognozavimo modelius ir juos palyginti tarpusavyje.

2. Juvenilinės hipertenzijos prognozavimo modeliai

Juvenilinės hipertenzijos prognozavimo modelių sudarymui pasirinkti logistinės regresinės analizės ir iki šiol tokio tipo tyrimams netaikytas neparimetrinės diskriminantinės analizės metodas [5, 7].



1 pav. Juvenilinės hipertenzijos tyrimo etapai.

Sudarant modelius panaudoti duomenys tik tų tiriamųjų, kurie dalyvavo visuose keturiuose tyrimuose (203 tiriamieji). Įvertinus medicinos ekspertų nuomonę, iš 158 stebėtų kintamųjų tolesnei analizei buvo atrinktas 21 kintamasis ir sudaryti 6 logistinės regresinės analizės modeliai (skirtingų lyčių 12–13 metų vaikams, 25–26 metų jaunuoliams ir 32–33 metų suaugusiems) skirti AH atsiradimo prognozavimui vyresniame amžiuje (38–39 metų), t.y., vertinant šių modelių parametrus nepriklausomų kintamųjų reikšmės buvo imamos iš pirmojo, antrojo arba trečiojo tyrimo, o priklausomojo kintamojo Y reikšmės (1 – žmogus serga AH, 0 – neserga AH) – iš paskutiniojo ketvirtojo tyrimo. Tikimybės, kad suaugęs 38–39 metų žmogus sirgs AH, įvertis (kai nepriklausomų kintamųjų vektorius yra $Z_j^{(i)}$) skaičiuojamas pagal formulę

$$\hat{P}(Y_j = 1 | Z_j^{(i)}) = \frac{e^{\hat{\Theta}^{(i)} Z_j^{(i)}}}{1 + e^{\hat{\Theta}^{(i)} Z_j^{(i)}}},$$

čia j yra tiriamojo numeris ($j = \overline{1, 203}$), i – tyrimo numeris ($i = \overline{1, 3}$), $\hat{\Theta}^{(i)} = (\hat{\theta}_0^{(i)}, \hat{\theta}_1^{(i)}, \dots, \hat{\theta}_p^{(i)})$ – logistinės regresijos modelių parametru įverčiai, p – nepriklausomų kintamųjų skaičius, $Z_j^{(i)} = (1, x_{j1}^{(i)}, x_{j2}^{(i)}, \dots, x_{jp}^{(i)})^T$ – j -ojo tiriamojo, i -ojo tyrimo rezultatai.

Parinkus logistinės regresijos modelį nustatyti kintamieji, kurie yra reikšmingi juvenilinės hipertenzijos prognozavimui. Gauta, kad visuose amžiaus grupėse (tyrimuose) reikšmingi yra šie kintamieji: arterinio kraujospūdžio lygiai ($X_1^{(i)}$ – diastolinis AKS ir $X_2^{(i)}$ – sistolinis AKS), antropometriniai duomenys ($X_3^{(i)}$ – ūgis, $X_4^{(i)}$ – svoris), $X_5^{(i)}$ – pulsas.

Prognozuojant AH atsiradimą vyresniame amžiuje 12–13 metų vaikams, be aukščiau paminėtų kintamųjų reikšmingi yra: $X_6^{(1)}$ – žasto apimtis, $X_7^{(1)}$ – odos raukšlės kairiojo tricepso srityje vidurkis, $X_8^{(1)}$ – odos raukšlės kairėje subskapuliarinėje srityje vidurkis, $X_9^{(1)}$ – krūtinės apimtis. 25–26 metų jaunuoliams be pirmųjų 5 kintamųjų juvenilinės hipertenzijos atsiradimo vyresniame amžiuje prognozavimui reikšmingi: $X_{10}^{(2)}$ – liemens apimtis, $X_{11}^{(2)}$ – klubų apimtis, $X_{12}^{(2)}$ – cholesterolis.

Naudojant tuos pačius atrinktus kintamuosius sudaryti juvenilinės hipertenzijos prognozavimo modeliai taikant neparametrinės diskriminantinės analizės metodą [5, 7]. Kiekvienos grupės anksčiau minėtų kintamųjų tankių įverčių radimui ir klasifikavimo kriterijų sudarymui buvo naudojami branduolio ir k -artimiausių kaimynų metodai [5]. Darbe panaudoti branduolių tipai pateikti 1 lentelėje.

Lentelėje V^{-1} yra atvirkštinė diskriminavimo kintamųjų kovariacijų matrica, r – spindulys, \mathbf{z} – p -matis vektorius, Γ – gama funkcija.

Iš viso sudaryta 18 modelių juvenilinės hipertenzijos prognozavimui (taikant logistinę regresinę analizę – 6 modeliai, taikant neparametrinės diskriminantinės analizės branduolio ir k -artimiausių kaimynų metodus – 12 modelių):

$$\text{LR}_i \left\{ \begin{matrix} v \\ m \end{matrix} \right\}, \quad \text{NDB}_i \left\{ \begin{matrix} v \\ m \end{matrix} \right\}, \quad \text{NDK}_i \left\{ \begin{matrix} v \\ m \end{matrix} \right\}, \quad i = \overline{1, 3},$$

čia LR – modelis sudarytas taikant logistinę regresinę analizę, NDB – taikant neparametrinės diskriminantinės analizės branduolio metodą, NDK – taikant neparametrinės diskriminantinės analizės k -artimiausių kaimynų metodą, v – vyrams, m – moterims, i – tyrimo, pagal kurio duomenis prognozuojama AH atsiradimas 38–39 metų amžiuje, numeris. Panaudojus sistemą SAS buvo sukurtos programos ir atlikta „mokymo“ imties klasifikacija prie taikytų metodų parametrų įvairių reikšmių (buvo keičiamas branduolio tipas ir spindulys r , kaimynų skaičius k). Kiekvienai klasifikacijai apskaičiuotas teisingo klasifikavimo procentas, klasifikavimo jautrumas ir specifišku-

1 lentelė. Branduolių tipai

Branduolys	Branduolio funkcija
Tolygusis	$K_t(\mathbf{z}) = \begin{cases} \frac{1}{v_r(t)}, & \text{jeigu } \mathbf{z}'\mathbf{V}_t^{-1}\mathbf{z} \leq r^2, \\ 0, & \text{kitu atveju,} \end{cases}$ $v_r(t) = r^p \mathbf{V}_t ^{1/2} v_0, \quad v_0 = \frac{\pi^{p/2}}{\Gamma(p/2+1)}.$
Normalusis	$K_t(\mathbf{z}) = [1/(c_0(t))] \exp(-[1/(2r^2)]\mathbf{z}'\mathbf{V}_t^{-1}\mathbf{z}),$ $c_0(t) = (2\pi)^{p/2} r^p \mathbf{V}_t ^{1/2}.$
Epanechnikovo	$K_t(\mathbf{z}) = \begin{cases} c_1(t) \left(1 - \frac{1}{r^2} \mathbf{z}'\mathbf{V}_t^{-1} \mathbf{z}\right), & \text{jeigu } \mathbf{z}'\mathbf{V}_t^{-1} \mathbf{z} \leq r^2, \\ 0, & \text{kitu atveju,} \end{cases}$ $c_1(t) = [1/(v_r(t))](1 + [p/2]).$

2 lentelė. Modelių palyginimo rezultatai

Modelis	Teisingai klasifikuota (%)	Jautrumas (%)	Specifiškumas (%)
LR1m	63,6	14,7	85,5
NDB1m (Epa, $r = 4$)	69,8	78,1	36,2
NDK1m ($k = 8$)	69,6	81,1	52,1
LR1v	55,9	51,1	60,9
NDB1v (Epa, $r = 5$)	65,8	60,9	75,9
NDK1v ($k = 7$)	72,1	70,0	74,4
LR2m	70,1	28,1	88,0
NDB2m (Epa, $r = 4$)	68,4	81,7	52,8
NDK2m ($k = 5$)	71,1	87,3	48,0
LR2v	55,2	58,5	47,7
NDB2v (Epa, $r = 7$)	60,0	55,9	70,0
NDK2v ($k = 5$)	70,3	70,8	68,9
LR3m	72,9	42,4	90,7
NDB3m (Epa, $r = 4$)	72,7	78,6	47,4
NDK3m ($k = 8$)	73,6	89,1	54,0
LR3v	62,8	56,1	68,9
NDB3v (Epa, $r = 7$)	63,6	60,6	86,6
NDK3v ($k = 8$)	67,7	77,8	63,8

mas. Geriausi „mokymo“ imties, t.y., tiriamųjų, klasifikavimo rezultatai pateikti 2 lentelėje. Matome, kad prognozuojant 12–13 metų vaikams ir 25–26 metų jaunuoliams AH atsiradimą vyresniame amžiuje ženkliai geresni rezultatai gauti taikant neparametrinės diskriminantinės analizės modelius. Lyginant su logistinės regresinės analizės modeliais teisingo klasifikavimo procentas padidėjo 16,2%, prognozuojant juvenilinę hipertenziją berniukams (12–13 metų), ir 15,1% – jaunuoliams (25–26 metų).

Žymėjimai 2 lentelėje: Epa – Epanechnikovo branduolys, r – spindulys, k – kaimynų skaičius.

Visų sudarytų neparametrinės diskriminantinės analizės modelių (2 lentelė) klasifikavimo kokybė buvo patikrinta naudojant Lachenbruch and Mickey procedūrą (drop-one cross-validation method) [7]. Klasifikavimo rezultatai suprastėjo tik 3–5%, todėl darome išvadą, kad panaudojus sukurtus modelius jau vaikystėje (12–13 metų) ir jaunystėje (25–26 metai) galima patenkinamai prognozuoti (teisingai klasifikuoja apie 70% atveju) AH atsiradimą vyresniame amžiuje (38–39 metų).

3. Išvados

1. Panaudojus logistinės regresinės analizės, neparametrinės diskriminantinės analizės metodus ir sistemą SAS sudaryti juvenilinės hipertenzijos atsiradimo prognozavimo modeliai, kurie leidžia patenkinamai (teisingai klasifikuoja 70–72%)

juvenilinės hipertenzijos atsiradimą 12–13 metų vaikams ir 25–26 metų jaunuoliams.

2. Palyginus logistinės regresinės ir neparametrinės diskriminantinės analizės modelių taikymo rezultatus nustatyta, kad neparametrinės diskriminantinės analizės modeliai geriau tinka juvenilinės hipertenzijos prognozavimui.

Literatūra

1. J. Klumbienė, Ž. Milašauskienė, I. Misevičienė, L. Šileikienė, A. Zaborskis, Ilgalaikis juvenilinės hipertenzijos tyrimas: arterinio kraujospūdžio kitimas nuo vaikystės iki jauno brandaus amžiaus, *Medicina*, **37**(12), 1526–1536 (2001).
2. L. Šileikienė, *Jaunų žmonių arterinės hipertenzijos rizikos veiksnių prognozinė reikšmė*, Dis. med. m. dr. laipsniui įgyti, KMU, Kaunas (2000).
3. W. Bao, S.R. Srinivasan, R. Valdez, K.J. Greenlund, W.A. Wattigney, G. Berenson, Longitudinal changes in cardiovascular risk from childhood to young adulthood in offspring of parents with coronary artery disease, *Jama*, **278**(21), 1452–1472 (1997).
4. J.-R. Kim, C.I. Kiefe, K. Liu, O.D. Williams, D.R. Jacobs, A. Oberman, Heart rate and subsequent blood pressure in young adults the Cardia study, *Hypertension*, **33**, 640–646 (1999).
5. J. Butler, Multivariate discrimination of binary data using SAS software, in: *Proceedings of 13th Annual Northeast SAS Users' Group Conference*, SAS Institute, Inc., Washington D.C. (2000).
6. W. Härdle, *Applied Nonparametric Regression*, Institut für Statistic und Ökonometrie, Berlin (1992).
7. *The SAS System OnlineDoc*, Cary, NC, USA, SAS Institute (2003).
8. E. Janavičiūtė, V. Janilionis, J. Klumbienė, Juvenilinės hipertenzijos prognozavimas naudojant neparametrinius diskriminantinės analizės metodus, in: *Matematika ir matematikos dėstymas*, Konferencijos pranešimų medžiaga, Technologija, Kaunas (2004), pp. 91–95.

SUMMARY

V. Janilionis, E. Janavičiūtė, J. Klumbienė, G. Sakalauskienė. Models for predicting juvenile hypertension

Aim of present study was to create some models to predict juvenile hypertension. Data was taken from longitudinal study of Juvenile hypertension in Lithuania. The first survey of a random sample of Kaunas schoolchildren born in 1964 was carried out in 1977. This population was re-examined in 1989, in 1997 and the last survey took place in 2002. There were created models for predicting juvenile hypertension using real data. The models were created using logistic regression, nonparametric discriminant methods (kernel and k -nearest neighborhood) and system SAS. The best results were received using nonparametric discriminant analysis.

Keywords: logistic regression, nonparametric discriminant analysis, juvenile hypertension, prediction.