Istraživanje čimbenika rizika za nastanak alergijskih dišnih bolesti s pomoću Bayesovih mreža temeljenih na podacima iz više izvora

Kovačić, Jelena

Doctoral thesis / Disertacija

2017

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, School of Medicine / Sveučilište u Zagrebu, Medicinski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:105:628040

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-28**



Repository / Repozitorij:

Dr Med - University of Zagreb School of Medicine Digital Repository





SVEUČILIŠTE U ZAGREBU
MEDICINSKI FAKULTET

Jelena Kovačić

Istraživanje čimbenika rizika za nastanak alergijskih dišnih bolesti s pomoću Bayesovih mreža temeljenih na podacima iz više izvora

DISERTACIJA

Zagreb, 2017.

SVEUČILIŠTE U ZAGREBU
MEDICINSKI FAKULTET

Jelena Kovačić

Istraživanje čimbenika rizika za nastanak alergijskih dišnih bolesti s pomoću Bayesovih mreža temeljenih na podacima iz više izvora

DISERTACIJA

Zagreb, 2017.

Disertacija je izrađena u Jedinici za medicinu rada i okoliša Instituta za medicinska istraživanja i medicinu rada u Zagrebu.

Voditelji rada: dr. sc. Veda Marija Varnai, dr. med., znan.savj. i prof. dr. sc. Anamarija Jazbec, dipl. ing. matematike

Sadržaj

1 Uvod i svrha rada 1

1.1 Alergijske dišne bolesti 1

1.2 Bayesove mreže i njihova primjena u epidemiologiji 4

1.3 Stapanje podataka iz više izvora 7

1.3.1 Procjena parametara iz djelomično preklapajućih podataka 9

1.3.2 Procjena strukture kauzalnih modela iz djelomično preklapajućih podataka 10

1.4 Svrha rada 10

2 Hipoteza 12

3 Ciljevi rada 12

4 Materijali i metode 13

4.1 Ispitanici 13

4.2 Metode 18

4.2.1 Dogradnja metode stapanja podataka iz više izvora 18

4.2.1.1 Procjena parametara modela G-MA metodom 18

4.2.1.2 Procjena parametara u meta-analizi 19

4.2.1.3 Procjena parametara u Bayesovim mrežama iz više izvora . 22

4.2.1.4 Evaluacija metode za procjenu parametara 22

4.2.1.5 Procjena strukture 28

4.2.1.6 Evaluacija metode za procjenu strukture 29
4.2.2 Otkrivanje rizičnih čimbenika alergijskih dišnih bolesti 34

5 Rezultati **37**

5.1 Dogradnja metode stapanja podataka iz više izvora **37**
5.1.1 Evaluacija metode G-MA za procjenu parametara 37
5.1.2 Evaluacija metode za procjenu strukture 41

5.2 Otkrivanje rizičnih čimbenika alergijskih dišnih bolesti **44**
5.2.1 Opće osobine ispitanika 44
5.2.2 Evaluacija metoda na testnom skupu 48
5.2.3 Rizični čimbenici za alergijske dišne bolesti 60
5.2.3.1 Rizični čimbenici za alergijske dišne bolesti: 1. izvor podataka 60
5.2.3.2 Rizični čimbenici za alergijske dišne bolesti: 2. izvor podataka 64
5.2.3.3 Rizični čimbenici za alergijske dišne bolesti: 3. izvor podataka 65
5.2.3.4 Zajednička analiza sva četiri izvora podataka 67
5.2.3.5 Povezanost kontakta sa životinjama s alergijskim dišnim bolestima 72
5.2.3.6 Povezanost pasivnog pušenja s alergijskim dišnim bolestima 73

6 Rasprava **73**

6.1 Dogradnja metode stapanja podataka iz više izvora **74**
6.1.1 Procjena parametara 75

6.1.2 Procjena strukture 77
6.1.3 Evaluacija na testnom skupu 79

6.2 Istraživanje čimbenika rizika za nastanak alergijskih dišnih bolesti 83

6.2.1 Povezanost alergijskih dišnih bolesti i alergijskih bolesti u djetinjstvu 84
6.2.2 Povezanost alergijskih dišnih bolesti i pozitivne obiteljske anamneze 86
6.2.3 Spolne razlike u prevalencijama alergijskih dišnih bolesti 88
6.2.4 Povezanost alergijskih dišnih bolesti i čimbenika životnog stila . . 90
6.2.4.1 Indeks tjelesne mase 90
6.2.4.2 Konzumacija alkohola 94
6.2.4.3 Izloženost duhanskom dimu 95
6.2.4.4 Tjelesna aktivnost 97
6.2.4.5 Povezanost depresije i alergijskih dišnih bolesti 99
6.2.5 Kontakt s mačkom 100
6.2.6 Centralno grijanje i alergijske dišne bolesti 101
6.2.7 Povezanost socioekonomskog položaja i alergijskih dišnih bolesti . 102
6.2.8 Povezanost neuroticizma i alergijskih dišnih bolesti 104
6.2.9 Povezanost prehrane s alergijskim dišnim bolestima 106
6.2.10 Dob i alergijske dišne bolesti 107
6.2.11 Povezanost mjesta stanovanja i alergijskih dišnih bolesti 109
6.2.12 Povezanost kvalitete života i alergijskih dišnih bolesti 111
6.2.13 Usporedba rezultata različitih vrsta analize 113
6.2.14 Zaključno 114

7 Zaključak **116**

8 Kratki sadržaj na hrvatskom jeziku **118**

9 Kratki sadržaj i naslov doktorskog rada na engleskom jeziku **119**

10 Popis literature **121**

11 Kratka biografija **145**

Popis oznaka i kratica

AS	aritmetička sredina
Bayes–G–MA	analiza Bayesovom mrežom pomoću G–MA metode
CC–MA	engl. *complete-case meta-analysis*
COmbINE	engl. *Causal discovery from Overlapping INtErventions*
G–MA	meta-analiza pomoću grafičkih modela
ITM	indeks tjelesne mase
MSE	engl. *mean-squared error*
NPV	negativna prediktivna vrijednost
PPV	pozitivna prediktivna vrijednost
ROC	engl. *receiver operating characteristic*
SD	standardna devijacija
WAO	engl. *World Allergy Organization*

1 Uvod i svrha rada

1.1 Alergijske dišne bolesti

Astma je kronična upalna bolest dišnih putova koja se klinički očituje ponavljajućim napadajima zviždanja u prsima, zaduhe, pritiska u prsima i kašlja uslijed opstrukcije dišnih putova koja je najčešće reverzibilna, bilo spontano ili uz liječenje (1). Rinitis se definira kao upalna bolest sluznice nosa koju karakteriziraju nosni simptomi curenja iz nosa, kihanja, začepljenosti nosa ili svrbeža nosa (2).

U alergijskoj astmi i alergijskom rinitisu upala je posljedica imunološke reakcije, najčešće posredovane IgE antitijelima, koja se u osjetljivih osoba javlja nakon dodira s alergenima iz okoline (2, 3). S obzirom da se naslijeđena sklonost organizma da stvara IgE protutijela kao odgovor na niske doze alergena koje se uobičajno pojavljuju u okolišu definira kao atopija (3), alergijska astma i alergijski rinitis posredovani IgE antitijelima često se nazivaju i atopijska astma i atopijski rinitis, iako se prema preporukama Svjetske alergološke organizacije (World Allergy Organization, WAO) kao točniji predlažu termini „IgE-posredovana alergijska astma" i „IgE-posredovan alergijski rinitis" (engl. *IgE-mediated allergic asthma*, *IgE-mediated allergic rhinitis*) (3).

Alergijske dišne bolesti od znatnog su javnozdravstvenog značaja. Procjenjuje se da od astme boluje oko 242 milijuna ljudi (procjena iz 2013. godine) (4), pri čemu je, uzevši u obzir da oko dvije trećine astmatičara ima alergijski tip astme (5), njih 161 milijuna bolovalo od alergijske astme. Pritom se prevalencija astme u svijetu u odnosu na 1990. godinu povećala za 32 % (4). Od alergijskog rinitisa, prema podacima Svjetske alergološke organizacije i Svjetske zdravstvene organizacije iz 2013. godine, boluje oko 400 milijuna ljudi: između 10 % i 30 % odraslih i do 40 % djece, pri čemu i prevalencija alergijskog rinitisa diljem svijeta nastavlja rasti (6).

Alergijske dišne bolesti predstavljaju znatan teret za pojedinca i društvo. Astma je 15. bolest u svijetu prema broju godina života koje pojedinac provede sa smanjenom sposobnosti zbog bolesti (engl. *years lived with disability*) (4). Također, bolest smanjuje učinak pojedinca u školi ili na poslu te negativno utječe na društveni život (7). Slično, umor, smanjena motivacija i koncentracija, tjeskoba, depresija, problemi sa spavanjem i općenito smanjena kvaliteta života zbog bolesti česti su u osoba s alergijskim rinitisom (6).

Gledano sa stajališta ekonomskog tereta bolesti za društvo, procjenjuje se da bi se samo na razini Europske Unije godišnje uštedjelo između 50 i 142 bilijuna eura kad bi radno aktivno stanovništvo koje boluje od alergijskih bolesti pravovremeno započelo s liječenjem (8).

Etiologija alergijskih dišnih bolesti još uvijek nije razjašnjena, a veže se uz složeni međuodnos genetskih i okolišnih čimbenika (9, 10, 11). Geni odgovorni za nastanak ovih bolesti slabe su ili nepotpune prodorne moći, te se znatna uloga pridaje i utjecaju okolišnih čimbenika. Brojni se okolišni čimbenici povezuju s nastankom alergijskih dišnih bolesti. Na primjer, u nedavno objavljenom preglednom radu autora Castro-Rodrigueza i suradnika (12) u kojem su skupljeni i analizirani rezultati meta-analiza u kojima su proučavani rizični i protektivni čimbenici za nastanak astme u djece, kao rizični negenetski čimbenici navode se, između ostalog, pretilost, psihološki stres i pušenje u majke tijekom trudnoće, izloženost pasivnom pušenju, prijevremeno rođenje i niska porođajna težina, porod Carskim rezom, teška infekcija respiratornim sincicijskim virusom rano u životu, prisutnost vlage i plijesni u kući, zagađenje zraka (prvenstveno dušikovim oksidima i sumporovim dioksidom), te pretilost i prekomjerna tjelesna masa, dok se kao protektivni čimbenici navode cijepljenje protiv tuberkuloze, prehrana bogata voćem i namirnicama s visokim sadržajem vitamina A i višestruko nezasićenih omega-3 masnih kiselina, smanjenje izloženosti alergenima u zraku i hrani, te život na selu i izloženost endo-

toksinu. Pri tome autori ističu da je kauzalna veza ustanovljena s visokim stupnjem sigurnosti samo za vrlo ograničeni broj čimbenika: prisutnost astme u roditelja kao nasljednog čimbenika, te nedonošenost i prenatalna izloženost duhanskom dimu kao negenetskih čimbenika. Za potvrdu kauzalne veze ostalih čimbenike potrebna su, prema autorima, daljnja istraživanja, intervencijska ili prospektivna (12).

Važnost okolišnih čimbenika u nastanku astme i alergijske senzibilizacije i u odraslih osoba pokazana je, primjerice, u istraživanju provedenim u osoba koje su migrirale iz područja s niskom u područja s višom prevalencijom alergijskih dišnih bolesti. Naime, u odraslih se migranata astma i alergijska senzibilizacija pojavila već nekoliko godina nakon doseljenja u područja s višom prevalencijom alergijskih bolesti (13). Istraživanje upućuje na moguću ulogu zagađenja zraka u industrijaliziranim područjima, te izloženost endotoksinu i„higijensku teoriju", prema kojoj smanjena izloženost mikrobima tijekom djetinjstva i adolescencije, kao i promjene crijevne mikroflore i učestala upotreba antibiotika, ometaju normalnu funkciju imunološkog sustava dajući prednost imunološkom odgovoru usmjerenom na bezazlene antigene, primjerice alergene (13).

Za alergijski rinitis vežu se slični rizični čimbenici kao i za astmu, primjerice poboljšanje higijenskih uvjeta života te posljedična manja izloženost bakterijskim i parazitarnim infekcijama („higijenska teorija"), izloženost dišnim i nutritivnim alergenima, urbani nasuprot ruralnom načinu života, zagađenje zraka cestovnim prometom i industrijom, porod Carskim rezom, nedonošenost, kemijska onečišćenja kućnog i radnog okoliša (npr. antimikrobnim sredstvima kao što su triklosan i parabeni), aktivno pušenje, socio-ekonomski status, slabije provjetravanje prostorija, prehrambeni čimbenici i sedentarni način života (14, 2, 15). Također se navodi obrnuta povezanost rizika s veličinom obitelji, te kontaktom s psom ili mačkom u ranom djetinjstvu (14). Kao i u slučaju astme, međutim, za većinu ovih čimbenika rezultati istraživanja nisu jednoznačni, što se i u slučaju alergijske astme i u slučaju

alergijskog rinitisa pokušava objasniti ne samo razlikama među ispitivanim populacijama ili dizajnu istraživanja, već i kompleksnim, još uvijek nedovoljno istraženim, međuodnosom genetskih i okolišnih čimbenika (14, 16, 17). Primjer takvog međuodnosa je štetni učinak pasivne izloženosti duhanskom dimu na težinu astme i razinu IgE antitijela samo u osoba s određenim polimorfizmima CD14 gena (gen uključen u imunološki odgovor na endotoksin) (18). I negenetski čimbenici mogu stajati u međuodnosu koji modificira njihov rizični ili zaštitni učinak. Primjerice, čini se da je utjecaj izloženosti alergenima na nastanak alergijskih dišnih bolesti ovisan o dobi. U istraživanju autorice Lynch i suradnika (19), izloženost alergenima grinja kućne prašine, glodavaca i žohara u djece starije od tri godine života povećala je rizik senzibilizacije, dok je izloženost u vrlo ranom djetinjstvu (tijekom prve godine života) imala zaštitni učinak.

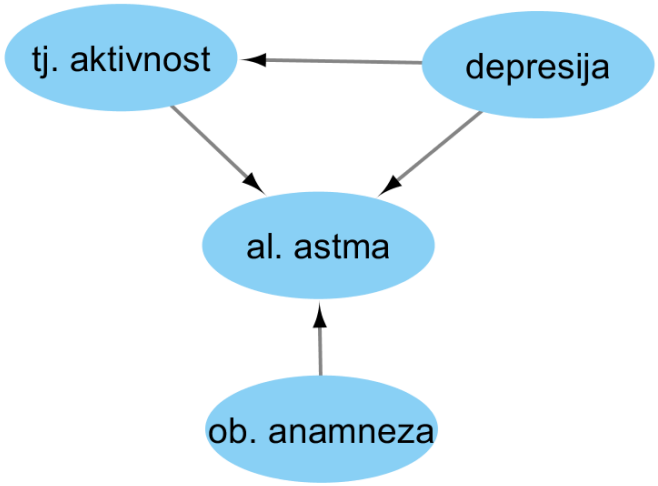
S obzirom na složene odnose među čimbenicima većinom malog do umjerenog utjecaja, kao sljedeći koraci koji bi mogli pomoći razjašnjenju etiologije alergijskih dišnih bolesti istaknuti su 1) primjena kompleksnijih, dosad malo korištenih statističkih modela u epidemiologiji respiratornih bolesti, poput Bayesovih mreža (20) te 2) stapanje podataka iz više raznorodnih izvora (11).

1.2 Bayesove mreže i njihova primjena u epidemiologiji

Bayesove mreže (21) vjerojatnosni su grafički modeli izdvojeni kao posebno prikladni za epidemiološka istraživanja (22). Prikazuju se usmjerenim acikličkim grafom (Slika 1) čiji čvorovi predstavljaju varijable (npr. rizične čimbenike i bolest), a usmjerene veze između čvorova interpretiraju se kao uzročno-posljedične.

Čvorovi i usmjerene veze među njima čine strukturu modela, dok jačine ovih veza čine parametre modela. Tako je izravna povezanost depresije i tjelesne aktivnosti,

predstavljena usmjerenom vezom između ove dvije varijable, dio strukture modela, dok je vjerojatnost da će se osoba redovito baviti tjelesnom aktivnošću ukoliko ima depresiju (broj u rasponu od 0 do 1, pri čemu viša vrijednost predstavlja veću vjerojatnost) jedan parametar modela. Ovakva se vjerojatnost još naziva i uvjetna vjerojatnost (engl. *conditional probability*). Svakom je čvoru pridružena uvjetna vjerojatnost uz dane čvorove-roditelje (čvorove usmjerene prema njemu). Tako, primjerice, čvoru za alergijsku astmu pripada uvjetna vjerojatnost uz danu tjelesnu aktivnost, depresiju i obiteljsku anamnezu (Slika 1). Parametri za ovaj čvor određeni su kad su određene vjerojatnosti prisutnosti bolesti za svaku kombinaciju vrijednosti ove tri varijable.

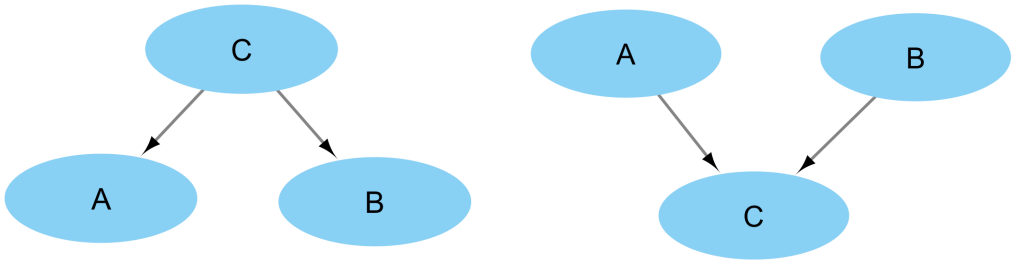


Slika 1: (a) Bayesova mreža prikazuje odnos alergijske astme, tjelesne aktivnosti, simptoma depresije i obiteljske anamneze. Slika je napravljena u programu Cytoscape.

Bayesove su mreže posebno prikladne za epidemiološka istraživanja jer omogućuju pravilno modeliranje odnosa među varijablama koje imaju zajednički uzrok (tzv. združivanje, engl. *confounding*) ili zajedničku posljedicu (Slika 2). Prikladni su i za modeliranje odnosa među visoko koreliranim čimbenicima, koji stvaraju probleme u regresijskoj analizi.

Nedostatak ovih modela je što ne mogu modelirati uzajamno djelovanje jedne

varijable na drugu: primjerice, hipoteza da depresija uzrokuje astmu, nakon čega astma pogoršava simptome postojeće depresije ne bi se mogla prikazati u ovakvom modelu.



(a) (b)

Slika 2: (a) Varijable A i B povezane su putem zajedničkog uzroka C (engl. *confounding*). (b) Varijable A i B povezane su putem zajedničke posljedice C. Slike su napravljene u programu Cytoscape.

I struktura i parametri mreže mogu se odrediti statističkim putem (računalnim algoritmima) ili koristeći samo vanjsko znanje (znanje stručnjaka i literaturno znanje), a moguća je i kombinacija ova dva pristupa.

Pregled računalnih algoritama za učenje strukture i parametara može se naći u literaturi (21). Nedostatak učenja strukture računalnim algoritmima je to što takvi algoritmi ne mogu zaista razlikovati uzrok od posljedice. Naime, u slučaju dvije povezane varijable jednaku potporu u podacima imaju hipoteze da je jedna varijabla 1) uzrok druge, 2) njena posljedica ili pak 3) povezana s njom putem nepoznatog zajedničkog uzroka (efekt združivanja). To treba imati na umu prilikom interpretacije tako dobivenog modela. Također, iz tog je razloga svakako poželjno koristiti vanjsko znanje (znanje stručnjaka ili literaturno znanje) u procesu izrade modela. Budući da Bayesove mreže na jednostavan i razumljiv način prikazuju uzročno-posljedične povezanosti između čimbenika i bolesti, stručnjaci iz odgovarajućeg područja medicine mogu se lako uključiti u ovaj proces.

U nedavnim epidemiološkim istraživanjima, Bayesove su mreže korištene u istraživanju odnosa između socioekonomskih pokazatelja i sub-kliničke ateroskleroze

(23), psiholoških čimbenika i prestanka pušenja (24), zaposlenja sa pretežno sjedilačkim načinom rada i bolest koronarnih arterija srca (25) te serumskog bilirubina i povišenog krvnog tlaka (26).

Ipak, u epidemiološkim su istraživanjima ovakvi modeli još uvijek rijetko korišteni u usporedbi s drugim metodama analize (poput regresijskih modela), pogotovo u području epidemiologije bolesti dišnog sustava (20). Jedan mogući razlog je taj što rezultati primjene modela na malim uzorcima mogu biti nepouzdani, što je još više izraženo kod bolesti niskih prevalencija (27).

Potreba za većim uzorcima u ovakvim istraživanjima jasna je. Ipak, epidemiološka istraživanja etiologije alergijskih dišnih bolesti su kompleksna, i u jednom zasebnom istraživanju neizvedivo je, iz etičkih, socijalnih, organizacijskih i financijskih razloga, u dovoljno velikom uzorku obuhvatiti sve rizične čimbenike za nastanak bolesti. Kao jedno moguće rješenje nameće se korištenje podataka iz više izvora u istoj analizi.

U nastavku je dan pregled trenutnih istraživanja i ograničenja metoda analize ovakvih podataka.

1.3 Stapanje podataka iz više izvora

Korištenje više raznorodnih izvora podataka u istoj analizi sve se više zagovara (28), pogotovo s obzirom da se očekuje sve više slobodno dostupnih podataka. Urednici časopisa PLOS Medicine proglasili su 2016. godinom kad slobodan pristup podacima postaje „univerzalni standard" u medicini (29). Međunarodni odbor urednika medicinskih časopisa (The International Committee of Medical Journal Editors, ICMJE) nedavno je pokrenuo inicijativu da podaci iz kliničkih pokusa budu dostupni u roku od šest mjeseci od objave rezultata pokusa (30), a slične se inicijative zagovaraju i za podatke iz epidemioloških istraživanja, uz argument da

je slobodan pristup podacima „ključno oružje u borbi protiv kroničnih nezaraznih bolesti" (31).

Potencijalna korist od upotrebe više različitih izvora podataka u istoj analizi velika je i sve se više prepoznaje. Dosadašnji radovi u području biomedicine, uključujući epidemiološka istraživanja, pokazali su dobrobit korištenja više izvora podataka u istoj analizi, npr. otkrivanje i smanjenje pristranosti (engl. *bias*) (32) i povećanje preciznosti procjene (33). Ipak, stapanje više izvora podataka nosi i brojna ograničenja.

Stapanje više izvora podataka u svakoj je primjeni problem za sebe. Izvori podataka mogu biti drastično različite kvalitete, pri čemu kod nekih može biti prisutna i pristranost, te je u svakoj primjeni nužno procijeniti u kojoj se mjeri i uz koja ograničenja izvori mogu stapati. Nadalje, definicije varijabli za isti čimbenik mogu se razlikovati od izvora do izvora (primjerice, bolest definirana temeljem različitih kriterija).

Statističke metode stapanja podataka su u razvoju, a velik je problem različita dostupnost varijabli u izvorima podataka (34). Primjerice, u različitim istraživanjima koja proučavaju istu bolest neki čimbenici utjecaja, kao što su dob i spol ispitanika, mogu biti dostupni u svim istraživanjima, dok su neki uključeni samo u dio istraživanja (npr. genska analiza). Ovakvi podaci nazivaju se djelomično preklapajući podaci (engl. *partially overlapping datasets*), a za čimbenike koji nisu uopće bili uključeni u istraživanje kaže se da sustavno nedostaju (engl. *systematically missing*).

1.3.1 Procjena parametara iz djelomično preklapajućih podataka

Metode za analizu ovakvih podataka tek su se unazad nekoliko godina počele razvijati u meta-analizi. Jedno predloženo rješenje bilo je višestruko umetanje nepotpunih podataka (engl. *multiple imputation*). Višestruko umetanje (35) standardno se koristi u nadopuni podataka kad je čimbenik uključen u istraživanje, ali za dio ispitanika podaci nedostaju (primjerice, dio ispitanika ne odgovori na pitanje ili ne sudjeluje u pretragama koje uključuju vađenje krvi).

Resche-Rigon i suradnici predložili su primjenu ove metode i za sustavno nepotpune podatke, ali samo za kontinuirane varijable (36), dok je za kategorijske varijable predloženo rješenje u radu autora Jolanija i suradnika (37). Drugo poboljšanje metode Resche-Rigona i suradnika predložili su nedavno Resche-Rigon i White (38), čija metoda osim sustavno nepotpunih čimbenika može nadopuniti i čimbenike koji su mjereni, ali s odazivom manjim od 100 %. Nedostatak svih ovih metoda je što u meta-analizu ne mogu uključiti sumarne podatke (primjerice, omjer šansi i njegovu standardnu pogrešku iz rezultata znanstvene publikacije), kakvi su česti u meta-analizama.

Metoda koja može uključiti i cijele izvore podataka i sumarne podatke iz publikacija predložena je u radu grupe Fibrinogen Studies Collaboration (39). Autori su predložili korištenje korelacija između procijenjenih koeficijenata iz različitih regresijskih modela, pri čemu se ovi modeli razlikuju prema grupama uključenih čimbenika (ovisno o dostupnosti čimbenika u istraživanju). Međutim, nedostatak metode je što u slučaju složenijih uzoraka sustavno nepotpunih podataka, kakvi su u stvarnosti česti, broj parametara koje treba procijeniti metodom postaje prevelik u odnosu na broj dostupnih izvora podataka. Zbog toga je jasna potreba za metodom koja bi smanjila broj parametara u takvim slučajevima, što bi je učinilo

primjenjivom u više stvarnih situacija s djelomično preklapajućim podacima.

1.3.2 Procjena strukture kauzalnih modela iz djelomično preklapajućih podataka

Dosad je razvijeno više metoda za učenje strukture kauzalnih modela iz više izvora djelomično preklapajućih podataka. Metode se razlikuju prema vrsti kauzalnih modela koje koriste te prema tome da li koriste podatke samo iz opažajnih istraživanja (engl. *observational studies*) ili mogu uključiti i eksperimentalna istraživanja, poput kliničkih pokusa. Pregled ovih metoda dan je u radu autora Triantafillou i Tsamardinos (40). Metoda COmbINE, predložena u radu ovih autora, može se koristiti za podatke iz opažajnih istraživanja, ali može uključiti i podatke iz eksperimentalnih istraživanja, pri čemu ima manje pretpostavki od ostalih takvih metoda te je stoga lakše primjenjiva u više stvarnih situacija. Metoda iz djelomično preklapajućih izvora podataka otkriva model koji je generalizacija Bayesove mreže (takozvani semi-Markovljev kauzalni model, engl. *Semi-Markov Causal Model*). U usporedbi s modelom Bayesove mreže, u ovom se modelu mogu modelirati i takozvane skrivene varijable (varijable koje nisu mjerene ni u jednom izvoru podataka).

Metoda COmbINE dosad je primijenjena samo u jednom primjeru stvarnih podataka (interakcija proteina u dvjema vrstama imunoloških stanica). Prema saznanjima autorice, metoda dosad nije upotrijebljena ni u jednoj epidemiološkoj primjeni, a nije korištena ni u kombinaciji s metodom za procjenu parametara.

1.4 Svrha rada

S obzirom na prethodno izložene probleme, svrha ovog istraživanja bila je dogradnja i poboljšanje postojeće metodologije za analizu podataka iz više izvora s preklapajućim čimbenicima te njena primjena u epidemiološkom istraživanju čimbenika rizika

za nastanak alergijskih dišnih bolesti iz četiri izvora odrasle populacije Hrvatske. Prema koracima u istraživanju, svrha rada je kako slijedi:

1. razvoj metode za procjenu parametara koja bi smanjila kompleksnost (u odnosu na broj parametara u procjeni) prethodno predložene metode za procjenu parametara djelomično preklapajućih podataka (39) (poboljšanje metodologije za procjenu parametara);

2. primjena metode razvijene u koraku 1 u izračunu parametara Bayesove mreže čija je struktura dobivena stapanjem djelomično preklapajućih podataka metodom COmbINE (dogradnja metode za procjenu strukture);

3. epidemiološka primjena metodologije dograđene u koraku 2 u prepoznavanju čimbenika povezanih s prisutnošću alergijskih dišnih bolesti i njihova međuodnosa (epidemiološka primjena).

2 Hipoteza

U istraživanju čimbenika rizika za nastanak alergijskih dišnih bolesti metoda analize djelomično preklapajućih podataka iz više izvora ima veću osjetljivost, specifičnost i pozitivnu i negativnu prediktivnu vrijednost u odnosu na analizu samo jednog izvora podataka.

3 Ciljevi rada

Opći cilj

Prepoznavanje čimbenika rizika za nastanak alergijskih dišnih bolesti, te njihovog međuodnosa, analizom djelomično preklapajućih podataka iz više izvora.

Specifični ciljevi

- Dogradnja metode stapanja podataka iz više izvora unutar generalizacije Bayesovih mreža uz procjenu mjera valjanosti (osjetljivost, specifičnost, pozitivna i negativna prediktivna vrijednost);
- Primjena dograđenog algoritma u istraživanju čimbenika rizika za nastanak alergijskih dišnih bolesti na podacima iz četiri epidemiološka izvora odrasle populacije Hrvatske, uz procjenu mjera valjanosti.

4 Materijali i metode

4.1 Ispitanici

U ovom radu korištena su četiri izvora podataka iz hrvatske populacije:

1. istraživanje „Alergotoksični učinci čimbenika općeg i radnog okoliša" (projekt MZOŠ br. 022-0222411-2410; 2007.-2014. godine); 439 studenata prve godine različitih preddiplomskih studija Sveučilišta u Zagrebu (41),

2. istraživanje „Imunotoksični učinci bioaerosola unutarnjeg okoliša i načina života" (projekt Instituta za medicinska istraživanja i medicinu rada - IMI, Zagreb; 2002.-2005. godine); 171 administrativni radnik (42),

3. podaci iz registra prethodnih pregleda obavljenih u Jedinici za medicinu rada i okoliša IMI-ja, od 1994. do 1998. godine, u svrhu zaposlenja u farmaceutskoj industriji 432 osobe različitih zanimanja i stručne spreme (43) i

4. podaci iz istraživanja World Health Survey (WHS)(44, 45) provedenog u Hrvatskoj tijekom 2003. godine u 993 odrasle osobe (pristup podacima može se zatražiti sa stranice (46)).

Prva tri izvora podataka sadrže informacije o okolišnim i nasljednim čimbenicima povezanim s nastankom atopijskih bolesti dišnih organa i kože, a četvrti izvor podatke o općem zdravstvenom stanju hrvatske populacije, uključujući podatke o prevalenciji alergijske astme, najčešćim zdravstvenim rizičnim čimbenicima i demografskim osobinama. U ovom istraživanju ispitivano je ukupno 18 čimbenika rizika, prikazanih u Tablici 1.

Od 18 ispitivanih čimbenika, samo su dva bila kontinuirane varijable (dob i indeks tjelesne mase), a svi ostali čimbenici imali su dvije moguće vrijednosti (kategorije). Kategorizacija ovih čimbenika opisana je u nastavku.

Tablica 1: Dostupnost čimbenika (prediktorskih varijabli) za svaki od epidemioloških izvora podataka. Mjereni čimbenici označeni su simbolom ✓.

Čimbenik	1. izvor	2. izvor	3. izvor	4. izvor
Spol	✓	✓	✓	✓
Dob	✓	✓	✓	✓
Pušenje	✓	✓	✓	✓
Indeks tjelesne mase	✓	✓	✓	✓
Stupanj obrazovanja	✓	✓	✓	✓
Život u gradu/selu	✓	✓	✓	✓
Veličina kućanstva	✓	✓		✓
Način grijanja u domu	✓	✓		✓
Alergijske bolesti u roditelja	✓	✓	✓	
Alergijske bolesti u djetinjstvu			✓	
Držanje kućnih ljubimaca	✓	✓	✓	
Tjelesna aktivnost	✓		✓	✓
Prehrambene navike	✓			✓
Pasivno pušenje	✓	✓		
Konzumacija alkohola				✓
Kvaliteta života		✓		✓
Neuroticizam	✓	✓		
Simptomi depresije	✓			✓

S obzirom na stupanj obrazovanja, ispitanici su podijeljeni u kategoriju s nižim stupnjem obrazovanja (srednjoškolsko obrazovanje i niže) te kategoriju s višim obrazovanjem (barem viša ili visoka stručna sprema). Mjesto boravišta (život u gradu ili u selu) u svim je izvorima bilo definirano kao trenutno mjesto stanovanja, osim u prvom izvoru gdje je označavalo mjesto stanovanja prije dolaska na fakultet. Veličina kućanstva bila je definirana prema broju soba po članu kućanstva. Kućanstva su podijeljena u kućanstva s manje od jedne sobe po osobi (prenapučena kućanstva, prema OECD-ovoj definiciji (47)) te kućanstva s barem jednom sobom po osobi. Budući da su, za razliku od prva dva izvora, u četvrtom izvoru u broj soba u kućanstvu uračunate i prostorije u kojima se ne boravi (primjerice, kupaona i kuhinja) (48), zbog neusporedivosti podataka informacija o veličini kućanstva u četvrtom izvoru isključena je iz daljnjih analiza.

Prema načinu grijanja, kućanstva su podijeljena u kućanstva s centralnim grijanjem (jedna kategorija) i ona s drugim vrstama grijanja (druga kategorija). Budući da se očekivao mali broj kućanstava koja se griju na plin, a nemaju centralno grijanje, te da postavljeno pitanje nije omogućavalo detaljnjiju razradu, u četvrtom izvoru u kućanstva s centralnim grijanjem uvrštena su sva kućanstva s plinskim grijanjem.

Pušenje je u svim izvorima definirano kao pušenje u vrijeme uzimanja upitnika. Pretjerana konzumacija alkohola definirana je prema smjernicama Svjetske zdravstvene organizacije (49, 50): muškarci koji su konzumirali barem pet standardnih alkoholnih pića te žene s konzumirana barem četiri pića tijekom barem jednog dana u tjednu koji je prethodio intervjuu označeni su kao neumjereni potrošači alkohola. Svi ostali ispitanici svrstani u kategoriju konzumacije alkohola koja je manje rizična za zdravlje.

Prehrambene navike bile su definirane u vidu (ne)dovoljnog unosa voća i povrća, prema smjernicama Svjetske zdravstvene organizacije za prevenciju kroničnih bolesti (51, 52). Ispitanici s dnevnim unosom voća i povrća nižim od 400 g bili su svrstani u kategoriju nedovoljnog unosa voća i povrća, a oni s unosom od barem 400 g u kategoriju dovoljnog unosa.

Tjelesna neaktivnost u prvom je i četvrtom izvoru definirana u skladu s preporukama Svjetske zdravstvene organizacije (53), na način opisan u radu (50). Kriterij dovoljne tjelesne aktivnosti zadovoljili su ispitanici koji tjedno provode barem 150 minuta u umjerenoj tjelesnoj aktivnosti ili 75 minuta u intenzivnoj aktivnosti, ili kombinacijom ovih aktivnosti (hodanje uključeno) postižu barem 600 MET-minuta tjedno (jedna MET-minuta iznosi 4184 $\mathrm{kJ{\cdot}kg^{-1}{\cdot}h^{-1}}$). Svi ostali ispitanici svrstani su u kategoriju nedovoljne tjelesne aktivnosti. U trećem izvoru kao nedovoljno aktivni označeni su ispitanici koji su odgovorili negativno na pitanje: „Bavite li se redovito sportom ili teško fizički radite i van rada (u slobodno vrijeme)”. Ispitanici iz trećeg

izvora koji su odgovorili pozitivno na ovo pitanje svrstani su u kategoriju dovoljne tjelesne aktivnosti.

Alergijske bolesti u roditelja zabilježene su ako je bilo tko od roditelja imao alergijsku bolest (astmu, peludnu hunjavicu ili ekcem). Prisutnost alergijskih bolesti u djetinjstvu izvedena je iz pitanja koje je obuhvaćalo ekcem, urtikariju, česte prehlade sa sviranjem u prsima i gušenjem te peludnu hunjavicu. Iako su i u prva dva izvora ispitanici pitani za alergijske bolesti ranije u životu, kod većine ispitanika zbog nepotpunih odgovora nije bilo moguće utvrditi da li se radilo o bolestima u djetinjstvu (do 10. godine života) ili kasnije.

Držanje kućnih ljubimaca u svim se izvorima odnosilo na redoviti kontakt sa životinjama u doba uzimanja intervjua, dok je u prvom izvoru obuhvaćen i period neposredno prije dolaska na fakultet. Kategorije ljubimaca bile su pas, mačka, ptica, glodavac i domaće životinje.

Kvaliteta života predstavljena je zadovoljstvom zdravljem iz WHOQOL-BREF upitnika za kvalitetu života (54). Ispitanici koji su na pitanje „Koliko ste zadovoljni svojim zdravljem" odgovorili da su nezadovoljni ili vrlo nezadovoljni svrstani su u jednu kategoriju, a ispitanici koji su odgovorili da su vrlo zadovoljni, zadovoljni ili niti zadovoljni niti nezadovoljni u drugu.

Neuroticizam je ispitan skalom neuroticizma Eysenckova upitnika ličnosti (55, 56) u prvom izvoru te upitnikom Cornell Index - Forma N3 (57, 42) u drugom izvoru. S obzirom da u oba upitnika veći rezultat označava višu razinu neuroticizma, ispitanici koji su imali rezultat veći od drugog tercila u uzorku (ispitanici čiji su rezultati bili u trećini najviših u uzorku) označeni su kao osobe s višom razinom neuroticizma.

Simptomi depresije u prvom su izvoru ispitani Zungovom skalom depresije (58, 56). Skala se sastoji od 20 pitanja – izjava vezanih za simptome depresije, a ispitanik odgovara na pitanje tako da odabere koliko se često osjeća onako kako je

opisano u danoj izjavi, u rasponu od „rijetko" do „većinu vremena". Rezultat na skali izražen je u rasponu od 25 do 100, pri čemu se rezultat od 60 i više smatra indikativnim za umjerenu ili težu depresiju. Za određivanje simptoma depresije u četvrtom izvoru korišten je nestandardni kriterij izveden iz sedam pitanja upitnika iz istraživanja World Health Survey Svjetske zdravstvene organizacije. Kriterij je izveden za potrebe istraživanja simptoma depresije u uzorku iz tog istraživanja i detaljno je opisan u drugim radovima (59, 60). Ukratko, ispitanici su bili svrstani u kategoriju ispitanika sa simptomima depresije ako su 1) odgovorili pozitivno na barem četiri od pet pitanja vezanih za simptome depresije u posljednjih 12 mjeseci te 2) odgovorili da su navedeni simptomi trajali većinu vremena u periodu dužem od dva tjedna.

Dostupnost čimbenika prema epidemiološkim izvorima vidljiva je iz Tablice 1. Šest čimbenika bilo je dostupno u sva četiri izvora podataka, pet čimbenika u tri izvora, pet u dva izvora, a preostala dva čimbenika bila su mjerena u samo jednom izvoru podataka.

Alergijske dišne bolesti definirane su na temelju subjektivnih i objektivnih pokazatelja atopije, u skladu s definicijom Svjetske alergološke organizacije (The World Allergy Organization, WAO-IAACI) (3). Subjektivni pokazatelji atopije, tj. simptomi astme i rinitisa, zabilježeni su tijekom usmene ankete provedene od strane liječnika. U prva 3 izvora alergijska astma definirana je kao prisutnost simptoma astme (zviždanje u prsima ili otežano disanje) (61) u posljednjih 12 mjeseci uz pozitivan kožni test ubodom (engl. *skin prick test*) na jedan ili više uobičajenih dišnih alergena (62) ili povišenu razinu ukupnih serumskih IgE protutijela (63), a alergijski rinitis kao prisutnost začepljenosti nosa ili najmanje dva od sljedećih nosnih simptoma: kihanje, svrbež nosa, iscjedak iz nosa (nepovezanih s akutnom infekcijom dišnih putova) (64) u posljednjih 12 mjeseci, uz pozitivan kožni test ubodom na jedan ili više uobičajenih dišnih alergena ili povišenu razinu ukupnih serumskih IgE protuti-

jela. Pritom su za obje alergijske bolesti vrijednosti IgE protutijela od 150 kU/L i više smatrane povišenima (64, 65, 66). U WHS izvoru alergijska astma definirana je na temelju dijagnoze postavljene od strane liječnika, uz prisutnost navedenih simptoma u posljednjih 12 mjeseci, dok prisutnost alergijskog rinitisa nije moguće utvrditi jer upitnik nije obuhvatio relevantna pitanja.

Kriterij za isključivanje iz kontrolne skupine bila je dijagnoza alergijske bolesti postavljena od strane liječnika (dostupna u drugom i trećem izvoru za dio ispitanika) te rad na radnom mjestu rizičnom za razvoj alergijskih dišnih bolesti duže od 6 mjeseci (podaci o radnoj povijesti poznati za većinu ispitanika iz drugog izvora).

4.2 Metode

4.2.1 Dogradnja metode stapanja podataka iz više izvora

4.2.1.1 Procjena parametara modela G-MA metodom

U sklopu ovog istraživanja prethodno je razvijena metoda za procjenu parametara kad podaci iz različitih istraživanja sadrže djelomično preklapajuće čimbenike. Metoda je objavljena u radu autorica Kovačić i Varnai (67) s detaljnim opisom i pojedinostima za implementaciju nužnim za primjenu metode.

Metoda je predložena za procjenu veličine učinka u regresijskim modelima (primjerice, za procjenu omjera šansi iz modela logističke regresije) i njegovog 95 %-tnog intervala pouzdanosti u meta-analizi, i u radu (67) evaluirana je na primjerima stvarnih podataka četiri epidemiološke meta-analize s nepotpunim podacima:

- 1. primjer: povezanost nesrazmjera truda i nagrade (engl. *effort-reward imbalance*) s ranim umirovljenjem (68),
- 2. primjer: povezanost razine fibrinogena s koronarnom srčanom bolesti (39),

- 3. primjer: povezanost debljine intime-medije karotidnih arterija i rizika oboljevanja od kardiovaskularnih bolesti (36),
- 4. primjer: povezanost alergijske astme i epizoda depresije (60).

U nastavku je dan kratak opis načina primjene metode i sažetak dobivenih rezultata te je opisano kako se metoda može primijeniti za procjenu parametara u Bayesovim mrežama iz više izvora.

4.2.1.2 Procjena parametara u meta-analizi

Primjer korištenja predložene metode autorica Kovačić i Varnai opisan je pomoću jednog epidemiološkog primjera na kojem je metoda evaluirana u radu (67).

Ovaj primjer dolazi iz rada (60) koji koristi podatke iz istraživanja World Health Survey (44, 45) Svjetske zdravstvene organizacije. Autori rada željeli su zaključiti postoji li povezanost između alergijske astme i simptoma depresije uz kontrolu za šest čimbenika: dob, spol, obrazovanje, tjelesna aktivnost, indeks tjelesne mase (ITM) i pušenje, to jest, željeli su procijeniti omjer šansi za alergijsku astmu iz jednadžbe logističke regresije oblika:

$$\begin{aligned}\ln\left(\frac{p}{1-p}\right) = \alpha_1 + \beta_1 \times \text{al. astma} + \gamma_{11} \times \text{spol} + \gamma_{12} \times \text{dob} + \gamma_{13} \times \text{obrazovanje} + \\ \gamma_{14} \times \text{ITM} + \gamma_{15} \times \text{pušenje} + \gamma_{16} \times \text{tj. aktivnost}\end{aligned} \tag{1}$$

Pritom p označava vjerojatnost prisutnosti simptoma depresije u promatranoj populaciji.

Budući da je u istraživanje bilo uključeno više različitih populacija (zemalja), u idealnom bismo slučaju procijenili koeficijent za alergijsku astmu (β_1) posebno za svaku promatranu populaciju te zatim izračunali sumarni omjer šansi pomoću standardnih metoda meta-analize. Međutim, s obzirom da su neke zemlje koristile skraćeni upitnik, samo je dio od 17 europskih zemalja imao svih šest dodatnih čimbe-

nika: jedna je zemlja imala podatke za sve čimbenike osim tjelesne aktivnosti, dok je jedna skupina zemalja imala podatke samo za spol, dob i obrazovanje te je stoga standardnim metodama bilo moguće izračunati samo omjer šansi korigiran za ova tri čimbenika.

Predloženom metodom bismo u ovom primjeru procijenili omjer šansi za alergijsku astmu iz jednadžbe (1) koristeći podatke iz svih populacija, uključujući i one koje imaju nepotpune podatke. Za primjenu metode, najprije je potrebno procijeniti regresijske koeficijente i njihove standardne pogreške, posebno za svako istraživanje u meta-analizi (u ovom primjeru, za svaku populaciju - zemlju), iz jednadžbe (1) te jednadžbi:

$$\ln\left(\frac{p}{1-p}\right) = \alpha_2 + \beta_2 \times \text{al. astma} + \gamma_{21} \times \text{spol} + \gamma_{22} \times \text{dob} + \gamma_{23} \times \text{obrazovanje} + \gamma_{24} \times \text{ITM} + \gamma_{25} \times \text{pušenje i} \quad (2)$$

$$\ln\left(\frac{p}{1-p}\right) = \alpha_3 + \beta_3 \times \text{al. astma} + \gamma_{31} \times \text{spol} + \gamma_{32} \times \text{dob} + \gamma_{33} \times \text{obrazovanje}. \quad (3)$$

Pritom je sva tri koeficijenta (β_1, β_2 i β_3) moguće procijeniti samo iz istraživanja s potpunim podacima (uključujući svih šest dodatnih čimbenika: spol, dob, obrazovanje, indeks tjelesne mase, pušenje i tjelesna aktivnost). Za zemlju čiji je upitnik sadržavao sve čimbenike osim tjelesne aktivnosti moguće je izračunati samo procjenu koeficijenata β_2 i β_3 iz jednadžbi (2) i (3), dok je za zemlje kod kojih su bili zabilježeni samo spol, dob i obrazovanje moguće procijeniti samo koeficijent β_3 iz jednadžbe (3).

Korelacije između procijenjenih koeficijenata $\hat{\beta}_1$, $\hat{\beta}_2$ i $\hat{\beta}_3$ iz iste populacije zatim se modeliraju u obliku Bayesove mreže. Naime, prema autorima ranije predložene metode (39), možemo očekivati da će koeficijenti $\hat{\beta}_1$, $\hat{\beta}_2$ i $\hat{\beta}_3$, procijenjeni iz iste populacije koristeći ove tri jednadžbe, biti korelirani te da možemo iskoristiti

njihovu korelaciju da smanjimo pristranost (to jest, da dobijemo točniju procjenu stvarnog efekta) i povećamo preciznost (to jest, da dobijemo uže intervale pouzdanosti) sumarnog omjera šansi u meta-analizi. U odnosu na ovu ranije predloženu metodu, procjena putem Bayesovih mreža predložena u radu autorica Kovačić i Varnai (67) može smanjiti broj parametara i pojednostavniti postupak njihove procjene, zahvaljujući tome što se u mrežama modeliraju samo veze među čvorovima koji su izravno povezani.

Tehnički detalji metodologije (način konstrukcije Bayesove mreže te način izračuna parametara), kao i rasprava o pretpostavkama modela, dani su u radu (67). S obzirom da metoda u procjeni koristi grafičke modele, u daljnjem je tekstu označena kraticom G-MA (meta-analiza pomoću grafičkih modela).

Metoda je u radu autorica Kovačić i Varnai (67) uspoređena sa standardnom metodom analize, meta-analizom samo onih istraživanja s potpunim podacima (engl. *complete-case meta-analysis*, u nastavku teksta označena kraticom CC-MA). Prema rezultatima iz rada, u primjeru meta-analize iz rada (68) omjer šansi dobiven G-MA metodom manje se razlikovao od omjera šansi dobivenog iz potpunih podataka (maksimalna razlika od 0,06) nego omjer šansi dobiven CC-MA metodom (maksimalna razlika od 0,31). Također, u osam primjera meta-analiza iz radova (68), (39) i (60), CC-MA metoda imala je 10 % do 64 % šire intervale pouzdanosti nego G-MA metoda. U jednom pak primjeru (primjer meta-analize iz rada (36)) rezultati dobiveni G-MA metodom nisu uzeti kao pouzdani zbog tehničkih problema u procjeni.

Metoda je primjenjiva i u slučaju kad istraživanja dolaze iz iste populacije (model s fiksnim efektima) i u slučaju istraživanja iz različitih populacija (model sa slučajnim efektima) te dopušta uključivanje i onih istraživanja za koje su poznati samo sumarni rezultati (kao što su objavljeni rezultati u obliku omjera šansi i njegove standardne pogreške). Iako je metoda primjenjiva u slučaju regresijskih modela bilo koje vrste, u ovom istraživanju korištena je samo procjena iz modela logističke

regresije, i to na način opisan u poglavlju 4.2.1.3.

4.2.1.3 Procjena parametara u Bayesovim mrežama iz više izvora

Metoda G-MA primjenjuje se za izračun jednog koeficijenta regresijskog modela u meta-analizi. Za dobivanje cijelog regresijskog modela, kojim bi u potpunosti bili određeni parametri u Bayesovim mrežama, metodu G-MA primijenili bismo zasebno za svaki čimbenik u modelu.

Iako postoji softver za modele s ovakvom parametrizacijom (primjerice, `R`-ov programski paket `abn` (69)), prema saznanjima autorice trenutno nema softvera u kojem bi se na jednostavan način dobili rezultati upita uvjetne vjerojatnosti (engl. *conditional probability query*) u ovakvim modelima, tj. u kojem bi se na jednostavan način mogla izračunati vjerojatnost bolesti za ispitanika za kojeg su poznate vrijednosti samo nekih čimbenika rizika.

Kako bi se izračunale vjerojatnosti bolesti, u ovoj je disertaciji iz tog razloga primijenjen pojednostavljen postupak. Iz dobivenih modela logističke regresije najprije su izračunate tablice uvjetne vjerojatnosti (engl. *conditional probability table*), uvrštavanjem vrijednosti čimbenika u dobiveni regresijski model. Iz ovih su tablica zatim primjenom naredbi `custom.fit` i `cpquery` `R`-ovog programskog paketa `bnlearn` dobivene željene vjerojatnosti.

4.2.1.4 Evaluacija metode za procjenu parametara

U poglavlju 4.2.1.3 opisano je kako se G-MA metoda autorica Kovačić i Varnai (67) može primijeniti za izračun veličine utjecaja čimbenika u Bayesovim mrežama iz više izvora. Naime, kako je već obrazloženo, procjena parametara se u tom slučaju svodi na procjenu regresijskih modela, pri čemu je zavisna varijabla uvijek određeni čvor, a prediktori njegovi roditelji. Iz tog je razloga evaluacija G-MA metode za procjenu parametara, opisana u ovom poglavlju, provedena na regresijskim modelima.

Budući da su rezultati evaluacije G-MA metode u slučaju kad izvori podataka dolaze iz različitih populacija (model meta-analize sa slučajnim efektima) prethodno objavljeni u radu autorica Kovačić i Varnai (67), nisu uključeni u ovu disertaciju u izvornom obliku, već su samo ukratko opisani zajedno s predloženom metodom u poglavlju 4.2.1.2.

U ovoj se disertaciji iz tog razloga detaljno opisuje samo evaluacija u slučaju kad izvori podataka dolaze iz iste populacije (model s fiksnim efektima). Evaluacija je provedena na umjetno stvorenim (računalno generiranim) podacima, kako je opisano u nastavku.

Generiranje podataka

Podaci su generirani pod pretpostavkom da su čimbenici koji utječu na pojavu bolesti isti te imaju jednaku veličinu efekta (tj. imaju jednak doprinos povećanju ili smanjenju rizika za pojavu bolesti) u svakom uzorku generiranom iz populacije, dok se prevalencije čimbenika mogu razlikovati od uzorka do uzorka. Iako su prave vrijednosti rizika za svaki uzorak jednake, njihove procijenjene vrijednosti iz različitih uzoraka mogu se, zbog statističkih pogrešaka koje nastaju zbog uporabe uzorka (engl. *sampling error*), bitno razlikovati.

Broj izvora podataka (3-4) te broj čimbenika motivirani su preliminarnim analizama četiri izvora podataka za alergijske bolesti iz hrvatske populacije. U preliminarnim analizama, čvorovi dobivenih grafičkih modela za alergijske bolesti imali su do tri čvora-roditelja te su iz tog razloga za evaluaciju odabrani modeli logističke regresije s tri čimbenika.

Provedena su ukupno tri eksperimenta, s 500 ponavljanja za svaki eksperiment. U svakom eksperimentu podaci za uzorke generirani su na isti način, no broj uzoraka i dostupnost mjerenih čimbenika po uzorku varirali su, kako je prikazano u Tablici 2. Svi generirani uzorci bili su veličine 500.

Tablica 2: Dizajn eksperimenata. Mjereni čimbenici označeni su simbolom ✓. Zavisna varijabla (prisutnost bolesti) mjerena je u svim izvorima podataka. (a) Eksperiment 1. (b) Eksperiment 2. (c) Eksperiment 3.

(a)

Uzorak	V_1	V_2	V_3
1	✓	✓	✓
2	✓		✓
3	✓		✓
4	✓		✓

(b)

Uzorak	V_1	V_2	V_3
1	✓	✓	✓
2	✓		✓
3	✓		✓
4	✓		

(c)

Uzorak	V_1	V_2	V_3
1	✓	✓	✓
2	✓		✓
3	✓		
4			

U svakom eksperimentu, podaci za zadani broj izvora podataka (uzoraka iz populacije) generirani su na sljedeći način. Za svaki izvor podataka, prvo su generirane vrijednosti za prediktorske varijable (čimbenike rizika i zaštitne čimbenike, to jest, čimbenike koji utječu na povećanje ili smanjenje rizika za bolest). Prediktorske varijable generirane su kao binarne, s mogućim vrijednostima 0 ili 1, iz binomne distribucije sa zadanom vjerojatnosti uspjeha u jednom pokusu. Vjerojatnost uspjeha, koja odražava prevalenciju generiranog čimbenika u populaciji, za svaki je izvor podataka i svaku varijablu odabrana u rasponu od 0,2 do 0,8 na slučajan način, iz uniformne distribucije sa zadanim rasponom.

Nakon generiranja prediktorskih varijabli (čimbenika), vjerojatnost p da je ispitanik bolestan izračunata je ovisno o prethodno generiranim vrijednostima prediktorskih varijabli za tog ispitanika, v_1, v_2 i v_3, na sljedeći način:

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3 \tag{4}$$

(standardni model logističke regresije) te je na temelju te vjerojatnosti ispitaniku dodijeljena vrijednost zavisne varijable 0 (odsustvo bolesti) ili 1 (prisutnost bolesti). Koeficijenti modela odabrani su na način da uključe jedan veći čimbenik rizika ($\beta_1 = \ln(2)$, što odgovara omjeru šansi 2) i jedan manji ($\beta_3 = \ln(1,1)$; odgovara omjeru šansi 1,1), te jedan zaštitni čimbenik ($\beta_2 = \ln(0,75)$, omjer šansi 0,75). Konstantni

član u jednadžbi (α) odabran je da odražava relativno nisku prevalenciju bolesti kad su svi drugi čimbenici odsutni, tj. imaju vrijednost 0 (koeficijent $\alpha = \ln(0,15)$, što odgovara prevalenciji od 13 %). Koeficijenti su u svakom eksperimentu bili fiksni kako bi se lakše ocijenile i usporedile značajke različitih metoda za procjenu parametara.

Nakon generiranja podataka, dio je podataka izbrisan na način prikazan u Tablici 2, a preostali su podaci korišteni u procjeni parametara.

Usporedba metoda za procjenu parametara

Metoda G-MA uspoređena je sa standardnim pristupom, logističkom regresijom samo jednog izvora podataka u kojem su mjereni svi čimbenici (metoda CC-MA). U svakom eksperimentu, G-MA metodom procijenjeni su koeficijenti α, β_1 i β_3 (ukupno tri grafička modela za svako ponavljanje eksperimenta). Koeficijent β_2 bilo je moguće procijeniti samo CC-MA metodom iz uzorka 1, u svakom ponavljanju, jer je varijabla V_2 bila mjerena samo u tom uzorku (Tablica 2) te je iz tog razloga ovaj koeficijent izuzet iz usporedbe.

Procijenjene vrijednosti za svaki koeficijent prikazane su srednjom vrijednosti i standardnom devijacijom, izračunatima za 500 ponavljanja eksperimenta. Raspodjela procijenjenih 500 vrijednosti prikazana je grafom gustoće. Također je, za svaki koeficijent, izračunata srednja vrijednost odstupanja procijenjenih vrijednosti od prave vrijednosti koeficijenta (pristranost, engl. *bias*) te srednja kvadratna pogreška (engl. *mean squared error*; MSE). Za svaku je metodu izračunata i pokrivenost (engl. *coverage*) koja predstavlja udio od 500 ponavljanja u kojima je procijenjeni interval pouzdanosti sadržavao pravu vrijednost koeficijenta. Uz pokrivenost, dana je i prosječna vrijednost omjera širina intervala pouzdanosti CC-MA i G-MA metode, izračunata također za 500 ponavljanja eksperimenta, za svaki koeficijent zasebno.

Implementacija G-MA metode

Kao i u radu autorica Kovačić i Varnai (67), za procjenu grafičkih modela u G-MA metodi korišten je program `WinBUGS` (70), statistički program koji parametre procjenjuje koristeći MCMC (engl. *Markov Chain Monte Carlo*) iteracije - računalno intenzivne simulacijske metode. Srednjim vrijednostima parametara dane su, isto kao i u navedenom radu, standardne neinformativne prvotne distribucije, $N(0, 1000^2)$. Početne vrijednosti parametara bile su srednje vrijednosti dostupnih koeficijenata. Vrijednosti standardnih devijacija koje su nedostajale uzorkovane su iz uniformne distribucije (u rasponu od 0 do trostruke vrijednosti najveće poznate standardne devijacije, na isti način kao u radu (67)).

Svi su rezultati temeljeni na završnih 50 000 MCMC iteracija, dobivenih s čimbenikom stanjivanja 100 nakon zadanog broja početnih iteracija. Budući da bi za ovako velik broj simulacija bilo vremenski previše zahtjevno detaljnim uvidom provjeriti konvergenciju svakog grafičkog modela koji je procijenjen G-MA metodom (ukupno 4500 grafičkih modela u sva tri eksperimenta), kao pokazatelj konvergencije korišten je rezultat Gewekeova testa (71). Ukoliko je hipoteza o konvergenciji odbačena za rezultate dobivene s početnih 1 000 000 iteracija, procjena istog modela ponovljena je s početnih 5 000 000 iteracija. Ukoliko je i tad hipoteza o konvergenciji odbačena, cijelo ponavljanje eksperimenta odbačeno je kao ponavljanje s neuspješnom konvergencijom.

Budući da procjena modela u rijetkim slučajevima može trajati i više od nekoliko sati, procjena je ograničena na 10 minuta za modele u eksperimentu 1 te na 20 minuta za modele u eksperimentima 2 i 3. Za ilustraciju, na 100 ponavljanja prvog eksperimenta 90 % modela procijenjeno je unutar jedne minute, za drugi eksperiment 90 % modela bilo je unutar dvije minute vremena procjene, a za treći eksperiment unutar minute i pol. Ponavljanja u kojima je za bilo koji od tri modela prekoračeno zadano vrijeme za procjenu odbačena su.

Zbog malog broja uzoraka, korelacijski koeficijenti nisu procjenjivani unutar modela, već su procijenjeni u zasebnoj simulacijskoj studiji prije provođenja eksperimenata. Podaci su u ovom slučaju generirani na isti način kao prije, s razlikom što veličine efekata čimbenika u populaciji α, β_1, β_2 i β_3 nisu bile fiksne, već su u svakom ponavljanju ove simulacijske studije za procjenu korelacija (ukupno 1000 ponavljanja) bile nasumično odabrane (α u rasponu od $\ln(0,1)$ do $\ln(0,2)$, a $\beta_1 - \beta_3$ u rasponu od $-\ln(3)$ do $\ln(3)$). Također, u svakom ponavljanju simulacijske studije generirano je ukupno 100 izvora podataka iz iste populacije kako bi se vjerodostojnije procijenili potrebni korelacijski koeficijenti.

Osim programa `WinBUGS`, korišten je i program `R`, verzija 3.1.1 (72) te `R` paketi `rube` (73) i `coda` (74).

4.2.1.5 Procjena strukture

Za procjenu strukture kauzalnih modela (uzročno-posljedičnih veza među varijablama) u ovom je istraživanju korištena metoda COmbINE autora Triantafillou i Tsamardinos (40). Metoda iz danih izvora podataka s preklapajućim čimbenicima nastoji otkriti kauzalni model koji je generirao sve dane podatke. Pritom se pretpostavlja da je taj model jednak za sve izvore podataka.

Kauzalni model koji metoda otkriva generalizacija je modela Bayesove mreže koja dopušta i modeliranje takozvanih „skrivenih" varijabli - varijabli koje nisu mjerene ni u jednom izvoru podataka. Tako je, osim standardne veze oblika A → B koja označava da A izravno uzrokuje (utječe na) B, u ovom modelu moguća i veza oblika A ↔ B, koja označava da postoji nepoznata (skrivena) varijabla koja izravno uzrokuje i A i B. Između dvije varijable istovremeno mogu postojati obje vrste povezanosti u modelu: A može izravno utjecati na B, a da u isto vrijeme postoji i skrivena varijabla koja utječe na obje ove varijable. Ipak, ova vrsta modela ne dopušta da uz vezu A → B postoji i veza B → A, bez obzira na to da li je prisutna veza A ↔ B ili ne.

Osim istraživanja s preklapajućim čimbenicima, unutar metode COmbINE mogu se modelirati i randomizirani kontrolirani eksperimenti (engl. *randomized controlled trial*) u kojima istraživač definira vrijednosti kontrolirane varijable: primjerice, varijable tretmana (lijek ili placebo) u istraživanju učinkovitosti lijeka.

Rezultat primjene metode je graf koji opisuje sve moguće kauzalne modele koji su mogli generirati podatke, takozvani sumarni graf (engl. *summary graph*). Ovaj sumarni graf može imati dvije vrste veza: veza koja nije isprekidana (—) označava da je veza sigurno prisutna[1], tj. nalazi se u svim kauzalnim modelima koji mogu objasniti generirane podatke, dok isprekidana veza (---) označava da je veza samo

[1]Ovdje se misli na vezu koju model matematički prepoznaje kao sigurno prisutnu, što ne mora značiti da je veza sigurno prisutna u stvarnom svijetu.

moguća (postoje kauzalni modeli koji mogu objasniti generirane podatke koji sadrže ovu vezu, kao i takvi modeli koji je ne sadrže). Ako veze između dvije varijable nema, znači da je nema ni u jednom kauzalnom modelu koji bi mogao generirati dane podatke.

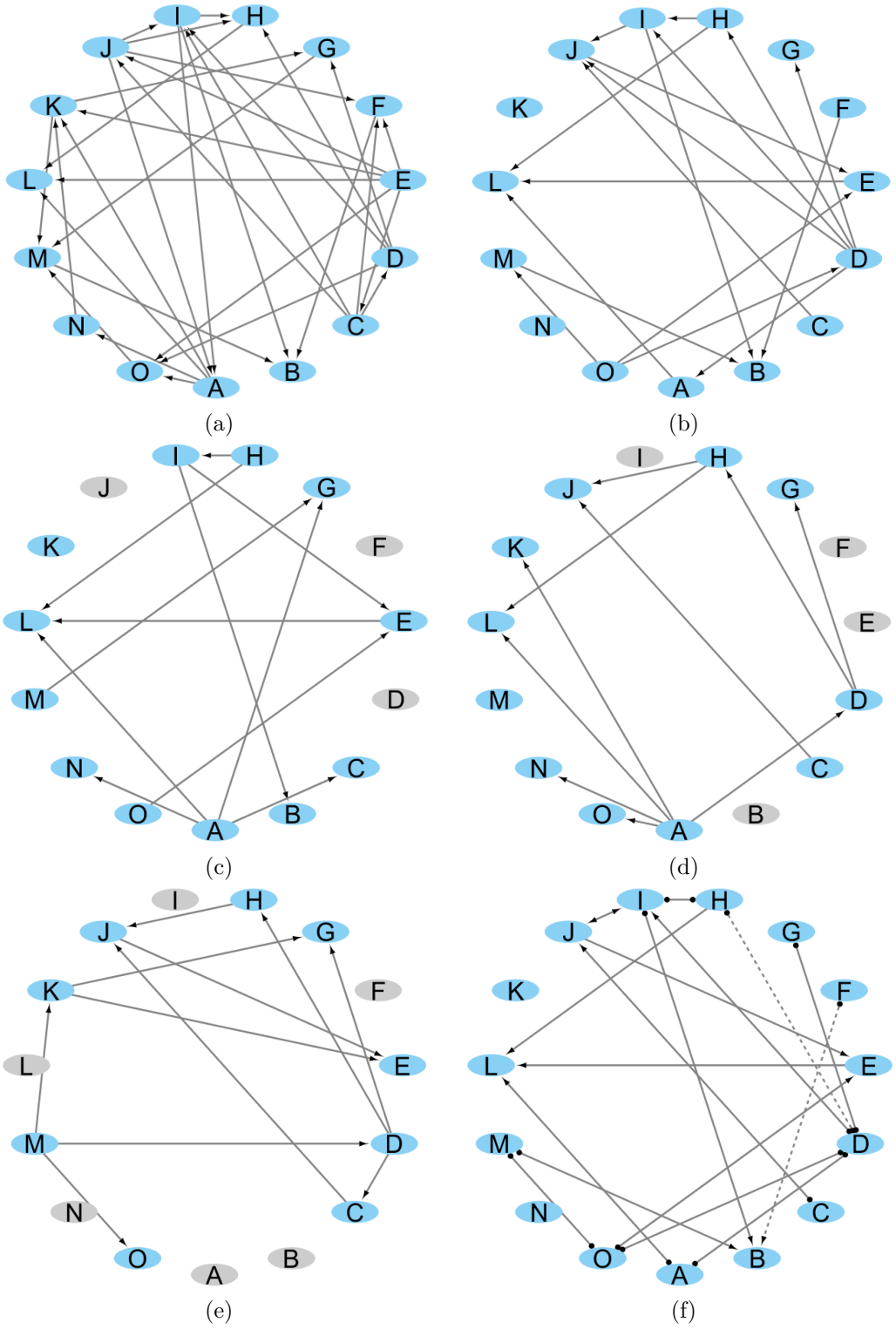
Isto tako, veza označena A → B u sumarnom grafu znači da je varijabla A označena kao izravan uzrok varijable B u svim mogućim kauzalnim modelima koji bi objasnili dane podatke, dok veze koje imaju različite orijentacije u različitim mogućim kauzalnim modelima počinju ili završavaju oznakom točke (•).

Detaljna teorija metode COmbINE i opsežni eksperimenti opisani su u radu autora metode (40). Za potrebe ovog istraživanja provedena je evaluacija koja je odgovarala epidemiološkoj primjeni iz ove disertacije, kako je opisano u nastavku.

4.2.1.6 Evaluacija metode za procjenu strukture

Za evaluaciju metode COmbINE odabrana su dva eksperimenta na umjetnim (računalno generiranim) podacima. U prvom su eksperimentu svi izvori podataka bili iz iste populacije (model s fiksnim efektima) te su, osim istih uzročno-posljedičnih veza, dijelili i njihovu jačinu. U drugom su eksperimentu generirani podaci iz različitih populacija (model sa slučajnim efektima). Iako su i u drugom eksperimentu podaci dijelili iste uzročno-posljedične veze, nisu nužno dijelili i jačine. Naime, njihove su jačine varirale među populacijama te je bilo moguće i da čimbenik u jednoj populaciji (tj. izvoru podataka) utječe na povećanje rizika (rizični čimbenik), a u drugoj na smanjenje (zaštitni čimbenik). Svaki eksperiment ponovljen je 100 puta. Primjer jednog ponavljanja eksprimenta dan je na Slici 3. Eksperimenti su provedeni na umjetnim (računalno generiranim) podacima.

Računalno generiranje podataka opisano je u nastavku.



Slika 3: Primjer jednog ponavljanja eksperimenta 1. (a) Originalni model. (b) - (e) Modeli dobiveni iz 4 izvora podataka. Sivom su bojom označene varijable izbrisane u danom izvoru. (f) Model dobiven primjenom metode COmbINE na modele (b)-(e). Veze koje nisu sigurne su iscrtane (--), dok veze s nesigurnim smjerovima počinju ili završavaju točkom (•). Slike su napravljene u programu Cytoscape.

Generiranje podataka

Za svako je ponavljanje eksperimenta najprije definiran kauzalni model, tj. uzročno-posljedične veze među varijablama (varijable uključuju čimbenike rizika, zaštitne čimbenike i prisutnost bolesti) u vidu Bayesove mreže (Slika 3 (a)). Bayesova je mreža u svakom ponavljanju izabrana na slučajan način, koristeći Ide-Cozmanov algoritam (75) implementiran u paketu `bnlearn` (76) programa `R` (72). Pritom je broj varijabli uvijek bio 15, a svaka varijabla imala je do tri čvora roditelja, tj. varijable o kojima direktno ovisi (na temelju preliminarnih analiza epidemioloških izvora podataka, opisanih u poglavlju 5.2.2).

Nakon što su za dano ponavljanje eksperimenta kauzalnim modelom određene uzročno-posljedične veze, jačine ovih veza određene su na sljedeći način. Za varijable koje su direktno ovisile o drugim varijablama (varijable sa čvorovima-roditeljima) jačine ovisnosti definirane su u regresijskom modelu sa čvorovima-roditeljima kao prediktorima. Koeficijenti regresijskih modela izabrani su na slučajan način, u rasponu od -1 do -0,25 ili od 0,25 do 1 za kontinuirane zavisne varijable (standardna linearna regresija) te u rasponu od $-\ln(2)$ do $-\ln(1,25)$ ili od $\ln(1,25)$ do $\ln(2)$ za binarne zavisne varijable (logistička regresija). Konstantni član u regresiji bio je 0 kad je zavisna varijabla bila kontinuirana, dok je za zavisne binarne varijable odabran u rasponu od $\ln(0,2)$ do $\ln(0,8)$. Varijable koje nisu direktno ovisile o drugim varijablama generirane su iz regresijskih modela sa samo konstantnim članom, bez prediktora. Varijable su podijeljene na kontinuirane i binarne na slučajan način prije određivanja jačina njihovih međuovisnosti.

U prvom su eksperimentu i kauzalni model i regresijski koeficijenti pomoću kojih su generirani podaci bili jednaki za sve generirane izvore podataka (model s fiksnim efektima). U drugom je eksperimentu kauzalni model bio jednak za sve izvore podataka, no koeficijenti su za svaki izvor bili drugačiji. Naime, nakon što je određeni koeficijent (npr., $\beta_1 = 0,5$) izabran na prethodno opisan način, za svaki

izvor podataka pripadni koeficijent izabran je iz normalne distribucije sa srednjom vrijednosti β_1 i varijancom 0,1 (standardni model sa slučajnim efektima). Zbog toga drugi eksperiment obuhvaća i slučajeve kad su u jednom izvoru dva čimbenika pozitivno povezana (npr. povećani rizik bolesti), a u drugom je povezanost slaba (neznatno izmijenjen rizik) ili negativna (smanjeni rizik).

Na temelju prethodno određenih veza i njihovih jačina, generirani su izvori podataka. U svakom ponavljanju generirana su četiri izvora podataka, s 500 ispitanika u svakom. Zatim je iz drugog, trećeg i četvrtog izvora namjerno izbrisan dio podataka. U drugom izvoru izbrisani su podaci za tri nasumce odabrane varijable (20 % podataka), u trećem za četiri (27 % podataka), a u četvrtom za šest (40 % podataka).

Nakon brisanja podataka, za svaki je izvor izgrađen zaseban kauzalni model Bayesove mreže (Slika 3 (b)–(e)). Pritom je korišten uzlazni algoritam (engl.*hill-climbing*) iz paketa `bnlearn` (76). Bayesova mreža izgrađena koristeći samo prvi izvor, s potpunim podacima, uspoređena je s modelom dobivenim pomoću algoritma COmbINE koristeći sva četiri modela s nepotpunim podacima (Slika 3 (f)). Ove dvije vrste modela uspoređene su s obzirom na broj točno prepoznatih uzročno-posljedičnih veza te njihovih smjerova (tj. točnog razlikovanja uzroka i posljedice).

Usporedba metoda za procjenu strukture

Budući da model Bayesove mreže na temelju podataka iz opažajnog istraživanja (engl. *observational study*) ne može razlikovati uzrok od posljedice (tj. u slučaju povezanih čimbenika A i B, modeli „A uzrokuje B" i „B uzrokuje A" imaju jednaku potporu u podacima, kako je objašnjeno u Uvodu), u usporedbi metoda za procjenu strukture zasebno su evaluirane točnost prepoznavanja veza (povezanosti) i točnost prepoznavanja smjerova (tj. razlikovanje uzroka od posljedice), kao i u radu autora

COmbINE metode (40).

Za točnost prepoznavanja veza korištene su mjere osjetljivosti, specifičnosti te pozitivne i negativne prediktivne vrijednosti. Pritom su stvarno pozitivne veze definirane kao veze koje su jednake (spajaju iste čvorove) u završnom modelu (Slika 3 (b) i (f)) i početnom kauzalnom modelu (Slika 3 (a)); lažno pozitivne veze kao one koje su nađene u završnom modelu, a nema ih u početnom; stvarno negativne veze kao one kojih nema ni u početnom ni u završnom modelu; lažno negativne one koje postoje u početnom modelu, ali ne i u završnom. Budući da CombINE metoda na temelju ulaznih podataka ne može uvijek odrediti da li veza mora postojati niti koji joj je smjer (Slika 3 (f)), u evaluaciji su u obzir uzete samo one veze koje su bile označene kao sigurno prisutne. U slučaju modela dobivenog samo iz prvog izvora s potpunim podacima (Slika 3 (b)) sve su veze bile sigurno prisutne.

Za točnost prepoznavanja smjerova korištene su mjere osjetljivosti i pozitivne prediktivne vrijednosti. Pozitivna prediktivna vrijednost za smjerove predstavljala je udio točnih smjerova veza u svim sigurnim smjerovima završnog modela. Osjetljivost za smjerove bila je udio svih smjerova u početnom kauzalnom modelu koji su točno prepoznati kao sigurni. Specifičnost i negativna prediktivna vrijednost u prepoznavanju smjerova nisu definirani zbog nemogućnosti definiranja stvarno negativnih smjerova.

Implementacija COmbINE metode

Implementacija COmbINE metode, dobivena od autora, preuzeta je s mrežne stranice (77). U evaluaciji je korišten program `MATLAB`, verzija R2016a (The MathWorks, Inc., Natick, Massachusetts, SAD).

4.2.2 Otkrivanje rizičnih čimbenika alergijskih dišnih bolesti

U ovom su dijelu disertacije opisane metode analize epidemioloških izvora podataka. Detaljan opis ovih izvora podataka nalazi se u poglavlju 4.1.

Kontinuirani podaci (dob i indeks tjelesne mase) prikazani su medijanom i interkvartilnim rasponom, dok su kategorijski podaci prikazani frekvencijama i relativnim frekvencijama ispitanika prema kategorijama. Za svaki je čimbenik izračunat i omjer šansi s 95 %-tnim intervalom pouzdanosti iz jednostruke logističke regresije, za svaki izvor podataka zasebno. U slučajevima kad omjer šansi nije bilo moguće pouzdano izračunati zbog nedovoljnog broja ispitanika u nekoj kategoriji, primijenjena je jednostruka Firthova logistička regresija (78).

Razmatrane su tri vrste modela za prepoznavanje čimbenika za nastanak alergijskih dišnih bolesti: višestruka logistička regresija i model Bayesove mreže dobiveni iz jednog izvora podataka te model Bayesove mreže dobiven iz sva četiri izvora. Pritom je, s obzirom da u četvrtom izvoru nisu bili poznati podaci za alergijski rinitis (pa tako niti za prisutnost alergijske dišne bolesti), ovaj izvor korišten samo za modeliranje međuodnosa čimbenika u izradi modela iz sva četiri izvora.

Bayesova mreža iz jednog izvora dobivena je uzlaznim algoritmom (engl. *hill climbing*) iz paketa `bnlearn` (76) statističkog programa `R` (72). U izgradnji mreža nisu bile dopuštene logički nemoguće direktne uzročno-posljedične veze. Tako nije bilo dopušteno da bilo koji atribut (osobina ispitanika ili značajka njegove okoline) utječe na atribut koji se pojavio u vremenu prije. Iz tog razloga niti jedan čimbenik osim stupnja obrazovanja, kao ni prisutnost alergijskih bolesti u ispitanika, nisu mogli utjecati na alergijske bolesti u roditelja (pokazatelj genskog naslijeđa dobivenog s rođenjem). S obzirom da stupanj obrazovanja ispitanika može biti pokazatelj socioekonomskog statusa cijele obitelji, uključujući roditelje,

ostavljena je mogućnost utjecaja stupnja obrazovanja ispitanika na alergijske bolesti u roditelja. Isto tako, onemogućeni su utjecaji svih atributa osim stupnja obrazovanja, alergijskih bolesti u roditelja i spola na alergijske bolesti ispitanika u djetinjstvu. Isključen je i utjecaj svih varijabli osim dobi, spola te alergijskih bolesti u roditelja i u djetinjstvu na stupanj obrazovanja (općenito stečen ranije u životu). Naposljetku, isključeni su utjecaji svih varijabli na dob i spol.

Mreže dobivene za svaki izvor zasebno integrirane su u jedan kauzalni model primjenom metode COmbINE (40). Parametri su zatim za svaki čvor u mreži procijenjeni novom metodom za procjenu parametara (G-MA) ako se dostupnost podataka za čvorove-roditelje razlikovala za izvore koji su imali mjeren odabrani čvor ili običnom meta-analizom (model s fiksnim efektima) u suprotnom. Ako je samo jedan izvor podataka sadržavao odabrani čvor, parametri su procijenjeni logističkom regresijom koristeći samo taj izvor.

U procjeni strukture modela dob i indeks tjelesne mase korišteni su kao kontinuirane varijable. Budući da je tablice uvjetne vjerojatnosti (engl. *conditional probability table*) moguće izračunati samo za kategorijske varijable, za procjenu parametara korištene su kategorizirane verzije ovih varijabli: prema dobi, ispitanici su podijeljeni u skupinu do 45 godina ($<$ 45 godina) i skupinu od 45 godina i više ($\geq$ 45 godina) (79), dok su prema indeksu tjelesne mase podijeljeni u one s prekomjernom tjelesnom težinom (indeks tjelesne mase $>$ 25) i ostale. Samo su dvije kategorije odabrane za svaku varijablu jer veći broj kategorija povećava broj parametara u tablici uvjetne vjerojatnosti i, s obzirom na veličine uzoraka, smanjuje pouzdanost procjene svakog parametra iz tablice.

Povezanost kontakta sa životinjama s alergijskim dišnim bolestima ispitana je zasebno samo kod ispitanika iz grada. Povezanosti su ispitane χ^2-testom ili, u slučaju kad je bilo premalo ispitanika za χ^2-test (u slučajevima kad je dvodimenzionalna kontigencijska tablica sadržavala ćeliju s manje od 5 ispitanika), Fisherovim

egzaktnim testom. Na isti je način ispitana povezanost pasivnog pušenja s alergijskim dišnim bolestima samo u nepušača.

Prije izgradnje konačnih modela, provedena je evaluacija modela na dijelu podataka na sljedeći način. Iz svakog izvora podataka izdvojen je dio podataka samo za evaluaciju modela (20 % nasumce odabranih podataka; u daljnjem tekstu skup za testiranje). Na preostalih 80 % podataka iz svakog izvora izgrađeni su modeli za evaluaciju: modeli logističke regresije i Bayesove mreže za svaki izvor zasebno te model Bayesove mreže za sva četiri izvora zajedno (dobiveni na prethodno opisan način). Ovi su modeli zatim uspoređeni na skupu za testiranje na temelju osjetljivosti, specifičnosti i prediktivnih vrijednosti.

Za modele Bayesove mreže vjerojatnost da ispitanik ima alergijsku bolest izračunata je za svaku jedinstvenu kombinaciju vrijednosti prediktora upitom uvjetne vjerojatnosti (engl. *conditional probability query*) s 50 000 000 ponavljanja, dok je za modele logističke regresije izračunata uvrštavanjem vrijednosti prediktora u dobivenu regresijsku jednadžbu. Budući da svi razmatrani modeli daju samo vjerojatnost da ispitanik ima alergijsku bolest te odluka modela da li je bolest prisutna ili ne ovisi o unaprijed određenoj graničnoj vrijednosti vjerojatnosti, odnos osjetljivosti i specifičnosti u ovisnosti o graničnoj vjerojatnosti (u rasponu od 0 do 0,99 s razmakom od 0,01) prikazan je ROC krivuljom. Za svaki je model izračunata i maksimalna postignuta točnost (udio točno klasificiranih ispitanika) i granične vjerojatnosti za koje se ta točnost postiže. Za te su vrijednosti granične vjerojatnosti izračunate pripadne pozitivne i negativne prediktivne vrijednosti.

Osim programa `R`, u opisanim analizama korišten je još i program `MATLAB R2016a` (The MathWorks, Inc., Natick, Massachusetts, SAD). Implementacija COmbINE metode, dobivena od autora, preuzeta je s mrežne stranice (77).

Slike grafičkih modela napravljene su u programu Cytoscape (80).

5 Rezultati

5.1 Dogradnja metode stapanja podataka iz više izvora

U ovom dijelu disertacije opisuju se rezultati vezani za razvoj i evaluaciju statističkih metoda. Primjena na četiri epidemiološka izvora (otkrivanje rizičnih čimbenika alergijskih dišnih bolesti) dana je u drugom dijelu rezultata.

5.1.1 Evaluacija metode G-MA za procjenu parametara

U ovom odjeljku sažeti su rezultati evaluacije metode G-MA za procjenu parametara u eksperimentima s računalno generiranim podacima[2]. Struktura generiranih podataka opisana je u poglavlju 4.2.1.4.

Korelacijski koeficijenti, potrebni u implementaciji G-MA metode, pokazali su visoke vrijednosti u zasebnoj simulacijskoj studiji: medijan dobivenih korelacija za koeficijente β_1 i β_3 kretao se od 0,95 do 1,00, a medijan za koeficijent α od 0,78 do 0,82. Iz tog su razloga u svim eksperimentima vrijednosti korelacija za β_1 i β_3 bile postavljene na 0,95, a za α na 0,8.

Rezultati eksperimenata prikazani su u Tablicama 3 – 5 i na Slikama 4 – 6.

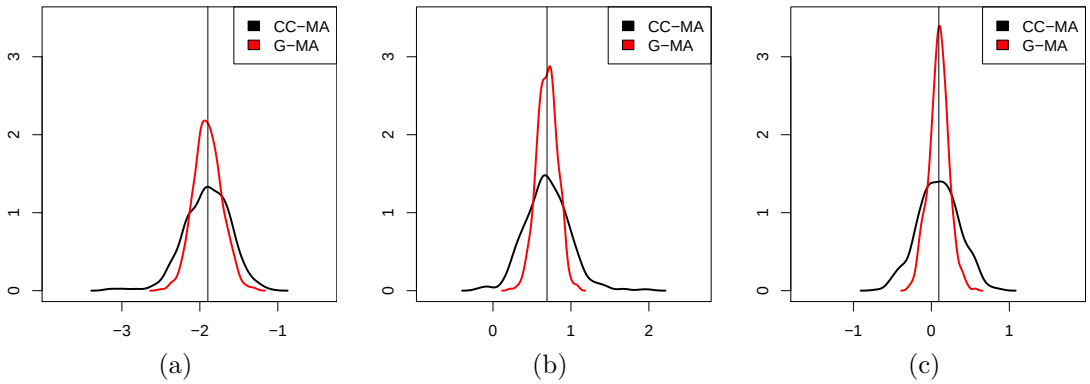
Na 500 uspješnih ponavljanja eksperimenta 1, bilo je 20 ponavljanja s neuspješnom kovergencijom (3,7 %) te 17 ponavljanja s prekoračenim zadanim vremenom izvođenja (3,2 %). U drugom eksperimentu, neuspješnih ponavljanja bilo je 28: 18 (3,4 %) s neuspješnom konvergencijom, 9 (1,7 %) s prekoračenim vremenom, te 1 ponavljanje (0,2 %) koje je završilo greškom treće vrste. Broj ponavljanja s neuspješnom konvergencijom u eksperimentu 3 bio je 30 (5,5 %), dok ih je 15 (2,8 %) prekinuto

[2]Rezultati su prikazani na skupu 21st Young Statisticians Meeting (4.– 6. studenog 2016., Piran, Slovenija) i uključeni su u obliku sažetka u zbornik sažetaka skupa (81).

zbog predugog izvođenja.

Raspodjela procijenjenih koeficijenata za 500 uspješnih ponavljanja po eksperimentu prikazana je na Slikama 4 – 6. Vidljivo je da su obje metode procjene imale sličnu razinu pristranosti, po apsolutnoj vrijednosti manju od 0,05 u svim eksperimentima (Tablice 3 – 5). Metode su bile slične i po pokrivenosti: u skoro svim slučajevima metode su imale pokrivenost jednaku ili veću od nominalne pokrivenosti od 95 %, tj. u skoro svim eksperimentima barem 95 % procijenjenih intervala pouzdanosti sadržavalo je pravu vrijednost procjenjivanog koeficijenta. Iznimka je bila samo CC-MA metoda u eksperimentu 1, gdje je 94 % procijenjenih intervala pouzdanosti za β_1 sadržavalo pravu vrijednost β_1.

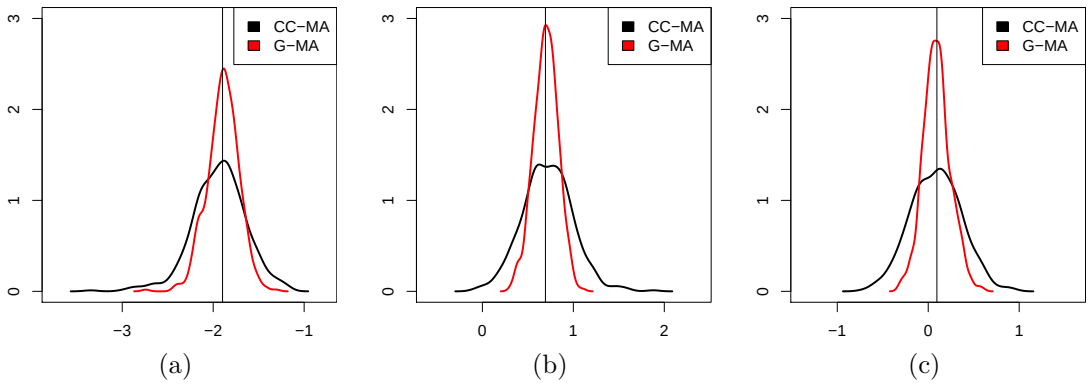
S druge strane, velika razlika među metodama bila je vidljiva kod usporedbe standardnih devijacija i srednje kvadratne pogreške. CC-MA metoda imala je konstantno veću standardnu devijaciju: 1,5 – 1,7 puta veću za koeficijent α, 1,8 – 2,2 puta veću za koeficijent β_1, te 1,4 – 2,1 puta veću u slučaju koeficijenta β_3. Razlika u srednjoj kvadratnoj pogrešci bila je još i veća: srednja kvadratna pogreška CC-MA metode bila je veća 2 – 3 puta od srednje kvadratne pogreške G-MA metode za koeficijent α, 3,5 – 4 puta za koeficijent β_1, te 2,3 – 4 puta za koeficijent β_3. Također, intervali pouzdanosti CC-MA metode bili su u svim slučajevima širi od intervala pouzdanosti G-MA metode (29 % – 39 % za koeficijent α, 56 % – 76 % za koeficijent β_1, te 36 % – 77 % za koeficijent β_3).



Slika 4: Raspodjele procijenjenih vrijednosti koeficijenata α (a), β_1 (b) i β_3 (c), dobivenih metodama G-MA i CC-MA, u eksperimentu 1. Vertikalna linija označava pravu vrijednost procjenjivanog koeficijenta.

Tablica 3: Rezultati usporedbe metoda CC-MA i G-MA za 500 ponavljanja u eksperimentu 1. AS: aritmetička sredina, SD: standardna devijacija, MSE: srednja kvadratna pogreška.

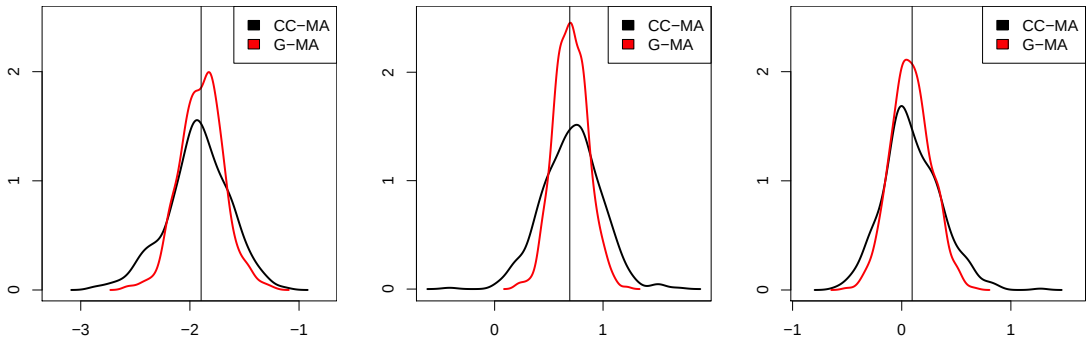
Koeficijent	Metoda	AS (SD)	Pristranost	MSE	Pokrivenost
α	CC-MA	$-1,91$ $(0,30)$	$-0,01$	$0,09$	$0,95$
$(-1,90)$	G-MA	$-1,90$ $(0,18)$	$-0,00$	$0,03$	$0,96$
β_1	CC-MA	$0,70$ $(0,29)$	$0,00$	$0,08$	$0,94$
$(0,69)$	G-MA	$0,69$ $(0,13)$	$0,00$	$0,02$	$0,97$
β_3	CC-MA	$0,09$ $(0,27)$	$-0,00$	$0,07$	$0,96$
$(0,10)$	G-MA	$0,09$ $(0,13)$	$-0,00$	$0,02$	$0,97$



Slika 5: Raspodjele procijenjenih vrijednosti koeficijenata α (a), β_1 (b) i β_3 (c), dobivenih metodama G-MA i CC-MA, u eksperimentu 2. Vertikalna linija označava pravu vrijednost procjenjivanog koeficijenta.

Tablica 4: Rezultati usporedbe metoda CC-MA i G-MA za 500 ponavljanja u eksperimentu 2. AS: aritmetička sredina, SD: standardna devijacija, MSE: srednja kvadratna pogreška.

Koeficijent	Metoda	AS (SD)	Pristranost	MSE	Pokrivenost
α	CC-MA	$-1,94$ $(0,30)$	$-0,04$	$0,09$	$0,95$
$(-1,90)$	G-MA	$-1,89$ $(0,18)$	$0,00$	$0,03$	$0,98$
β_1	CC-MA	$0,72$ $(0,28)$	$0,02$	$0,08$	$0,95$
$(0,69)$	G-MA	$0,70$ $(0,13)$	$0,00$	$0,02$	$0,97$
β_3	CC-MA	$0,09$ $(0,28)$	$-0,01$	$0,08$	$0,95$
$(0,10)$	G-MA	$0,09$ $(0,15)$	$-0,00$	$0,02$	$0,97$



(a) (b) (c)

Slika 6: Raspodjele procijenjenih vrijednosti koeficijenata α (a), β_1 (b) i β_3 (c), dobivenih metodama G-MA i CC-MA, u eksperimentu 3. Vertikalna linija označava pravu vrijednost procjenjivanog koeficijenta.

Tablica 5: Rezultati usporedbe metoda CC-MA i G-MA za 500 ponavljanja u eksperimentu 3. AS: aritmetička sredina, SD: standardna devijacija, MSE: srednja kvadratna pogreška.

Koeficijent	Metoda	AS (SD)	Pristranost	MSE	Pokrivenost
α	CC-MA	$-1,93$ $(0,29)$	$-0,03$	$0,08$	$0,95$
$(-1,90)$	G-MA	$-1,89$ $(0,20)$	$0,00$	$0,04$	$0,97$
β_1	CC-MA	$0,71$ $(0,27)$	$0,02$	$0,07$	$0,96$
$(0,69)$	G-MA	$0,70$ $(0,15)$	$0,01$	$0,02$	$0,97$
β_3	CC-MA	$0,10$ $(0,26)$	$-0,00$	$0,07$	$0,97$
$(0,10)$	G-MA	$0,08$ $(0,18)$	$-0,01$	$0,03$	$0,96$

5.1.2 Evaluacija metode za procjenu strukture

U ovom potpoglavlju sažeti su rezultati evaluacije metode COmbINE za procjenu strukure Bayesove mreže iz više izvora podataka. Generiranje podataka i provedeni eksperimenti opisani su u potpoglavlju 4.2.1.6.

Rezultati prvog eksperimenta prikazani su u Tablicama 6 i 7. Bez obzira na broj korištenih izvora, većina prepoznatih direktnih povezanosti (veza u dobivenom grafu) bila je točna (medijani pozitivnih prediktivnih vrijednosti od 0,84 u slučaju modela dobivenog iz jednog izvora podataka te 0,86 u slučaju modela četiri izvora podataka; Tablica 6). Također, u većini slučajeva gdje modeli nisu našli direktnu uzročno-posljedičnu vezu takve veze nije ni bilo (medijani negativnih prediktivnih vrijednosti od 0,84 za model jednog izvora podataka te 0,83 za četiri izvora podataka). Ipak, iako su oba modela vrlo točno prepoznavala slučajeve gdje doista nema direktne povezanosti (medijani specifičnosti od barem 0,95), dio direktnih uzročno-posljedičnih veza ostao je neprepoznat. Kod modela jednog izvora podataka, medijan osjetljivosti bio je 0,60. Usprkos sličnoj pozitivnoj prediktivnoj vrijednosti, kod modela sa četiri izvora podataka osjetljivost je bila i niža (medijan 0,50) zbog povezanosti koje su bile klasificirane samo kao moguće.

Tablica 6: Rezultati eksperimenta 1 prema otkrivenim povezanostima, u obliku: medijan (2,5 % - 97,5 %-tni kvantil) za 100 ponavljanja. PPV: pozitivna prediktivna vrijednost, NPV: negativna prediktivna vrijednost.

Korišteni izvori	PPV	NPV	Osjetljivost	Specifičnost
1 izvor	$0,84$ $(0,60-1,00)$	$0,84$ $(0,74-0,96)$	$0,61$ $(0,26-0,92)$	$0,95$ $(0,76-1,00)$
4 izvora (COmbINE)	$0,86$ $(0,60-1,00)$	$0,83$ $(0,73-0,93)$	$0,50$ $(0,21-0,78)$	$0,96$ $(0,81-1,00)$

Oba su modela lošije prepoznavala smjerove povezanosti (Tablica 7). Tako je za oba modela u prosjeku tek oko 50 % prepoznatih povezanosti bilo ispravno ori-

jentirano, tj. točno je razlučen uzrok od posljedice (medijani pozitivnih prediktivnih vrijednosti 0,46). U kombinaciji s nižim vrijednostima osjetljivosti kod prepoznavanja povezanosti za model s jednim izvorom, u konačnici je za u prosjeku malo više od trećinu direktnih uzročno-posljedičnih veza točno utvrđena i povezanost i njezin smjer (medijan osjetljivosti 0,34). Osjetljivost kod prepoznavanja smjerova povezanosti za model sa četiri izvora podataka bila je još niža (medijan 0,17), opet zbog dijela povezanosti i smjerova koje model nije mogao sa sigurnošću potvrditi.

Tablica 7: Rezultati eksperimenta 1 prema smjerovima otkrivenih povezanosti, u obliku: medijan (2,5 % - 97,5 %-tni kvantil) za 100 ponavljanja. PPV: pozitivna prediktivna vrijednost.

Korišteni izvori	PPV	Osjetljivost
1 izvor	$0,46$ $(0,18-0,75)$	$0,34$ $(0,10-0,74)$
4 izvora (COmbINE)	$0,46$ $(0,02-1,00)$	$0,17$ $(0,00-0,53)$

Tablica 8: Rezultati eksperimenta 2 prema otkrivenim povezanostima, u obliku: medijan (2,5 % - 97,5 %-tni kvantil) za 100 ponavljanja. PPV: pozitivna prediktivna vrijednost, NPV: negativna prediktivna vrijednost.

Korišteni izvori	PPV	NPV	Osjetljivost	Specifičnost
1 izvor	$0,80$ $(0,60-1,00)$	$0,82$ $(0,76-0,95)$	$0,56$ $(0,28-0,89)$	$0,93$ $(0,81-1,00)$
4 izvora (COmbINE)	$0,81$ $(0,56-1,00)$	$0,81$ $(0,74-0,91)$	$0,47$ $(0,24-0,72)$	$0,94$ $(0,84-1,00)$

U drugom eksperimentu (podaci iz različitih populacija) oba su se modela ponašala vrlo slično kao u prvom eksperimentu, iako su vrijednosti mjera točnosti bile općenito malo niže (Tablice 8 i 9). Opet je većina odnosa među varijablama u dobivenim modelima točno klasificirana kao prisustvo, odnosno odsustvo, direktne uzročno posljedične veze (medijani prediktivnih vrijednosti 0,80 i više). Specifičnost

modela i dalje je bila visoka (medijani 0,93 - 0,94), dok je osjetljivost bila niža (medijan od 0,56 za model s jednim izvorom podataka te 0,47 za četiri izvora podataka).

Slično kao i prije, za manji dio prepoznatih veza točno je određen uzrok i posljedica (medijani pozitivnih prediktivnih vrijednosti 0,47 i 0,45; Tablica 9), što se opet odrazilo na osjetljivosti kod prepoznavanja smjerova povezanosti (medijan 0,32 za model s jednim izvorom te 0,19 za model sa četiri izvora podataka).

Tablica 9: Rezultati eksperimenta 2 prema smjerovima otkrivenih povezanosti, u obliku: medijan (2,5 % - 97,5 %-tni kvantil) za 100 ponavljanja. PPV: pozitivna prediktivna vrijednost.

Korišteni izvori	PPV	Osjetljivost
1 izvor	$0,47$ $(0,24-0,71)$	$0,32$ $(0,14-0,66)$
4 izvora (COmbINE)	$0,45$ $(0,00-1,00)$	$0,19$ $(0,00-0,47)$

5.2 Otkrivanje rizičnih čimbenika alergijskih dišnih bolesti

5.2.1 Opće osobine ispitanika

Deskriptivna statistika za svaki od epidemioloških izvora podataka prikazana je u Tablici 10.

Iz kontrolne skupine drugog izvora isključeno je šest ispitanika s postavljenom dijagnozom alergijske dišne bolesti od strane liječnika te je konačan broj ispitanika u tom izvoru bio 165. Iz trećeg je izvora iz istog razloga isključen jedan ispitanik. Dodatno su iz trećeg izvora isključena i 92 ispitanika koji su radili duže od šest mjeseci na radnom mjestu rizičnom za razvoj alergijskih dišnih bolesti i 30 ispitanika s nepotpunim podacima o radnom mjestu te je konačan broj ispitanika u ovom izvoru bio 309. Iz prvog i četvrtog izvora nije isključen nijedan ispitanik te su konačne veličine uzorka za ova dva izvora bile redom 439 i 993.

U prvom izvoru bilo je 96 (22 %) ispitanika s alergijskom dišnom bolesti, od kojih je 24 imalo i alergijsku astmu i alergijski rinitis, 8 ih je imalo samo alergijsku astmu, a 64 samo alergijski rinitis. U drugom izvoru bilo je 46 (28 %) ispitanika s alergijskom dišnom bolesti, od kojih je 17 imalo i alergijsku astmu i rinitis. Još je dvoje ispitanika imalo samo alergijsku astmu te 27 samo alergijski rinitis. Od 21 (7 %) ispitanika s alergijskim dišnim bolestima u trećem izvoru, 20 ih je imalo alergijski rinitis, a jedan je imao alergijsku astmu. U četvrtom izvoru broj ispitanika s alergijskom astmom bio je 51, dok za alergijski rinitis podaci nisu bili poznati.

U prva dva izvora više od dvije trećine ispitanika bile su žene (udjeli od 72 % i 79 %). Žene su bile većina i u četvrtom izvoru podataka, iako s manjim udjelom (58 %), dok je u trećem izvoru bilo nešto više muškaraca (54 %). Prema dobi, najmlađi su ispitanici bili u prvom i trećem izvoru (medijan dobi 19 i 24 godine), dok je u

Tablica 10: Deskriptivna statistika za epidemiološke izvore podataka. N označava broj ispitanika s poznatim vrijednostima. Udio ispitanika koji pasivno puše prikazan je samo za nepušače.

Čimbenik	1. izvor $N: 439$	2. izvor $N: 165$	3. izvor $N: 309$	4. izvor $N: 993$
Al. dišna bolest	96 (22 %) $N: 432$	46 (28 %) $N: 165$	21 (7 %) $N: 306$	
Muški spol	123 (28 %) $N: 439$	34 (21 %) $N: 165$	166 (54 %) $N: 309$	414 (42 %) $N: 990$
Dob	19 (18 - 19) $N: 439$	43 (37 - 51) $N: 165$	24 (21 - 27) $N: 309$	50 (37 - 62) $N: 990$
Pušenje	153 (35 %) $N: 439$	57 (35 %) $N: 165$	129 (42 %) $N: 309$	262 (27 %) $N: 985$
Indeks tjelesne mase	21 (20 - 23) $N: 439$	24 (22 - 27) $N: 165$	22 (21 - 24) $N: 301$	26 (23 - 29) $N: 980$
VSS/VŠS obrazovanje		81 (49 %) $N: 165$	90 (29 %) $N: 309$	82 (8 %) $N: 990$
Život u gradu	328 (75 %) $N: 438$	151 (92 %) $N: 165$	245 (80 %) $N: 306$	668 (67 %) $N: 993$
Bar 1 soba po osobi	239 (55 %) $N: 432$	99 (60 %) $N: 165$		
Centralno grijanje	291 (67 %) $N: 437$	141 (85 %) $N: 165$		299 (31 %) $N: 974$
Al. bolesti u roditelja	109 (25 %) $N: 437$	19 (12 %) $N: 163$	28 (9 %) $N: 303$	
Al. bolesti u djetinjstvu			20 (7 %) $N: 295$	
Kućni ljubimci:				
Pas	213 (49 %) $N: 437$	39 (24 %) $N: 163$	84 (29 %) $N: 291$	
Mačka	156 (36 %) $N: 436$	41 (25 %) $N: 163$	66 (23 %) $N: 291$	
Ptica	24 (5 %) $N: 438$	8 (5 %) $N: 164$	35 (12 %) $N: 291$	
Glodavac	29 (7 %) $N: 438$	9 (6 %) $N: 163$	5 (2 %) $N: 291$	
Domaće životinje	47 (11%) $N: 439$	5 (3 %) $N: 164$	37 (13 %) $N: 291$	
Dovoljna tj. aktivnost	412 (94 %) $N: 439$		80 (28 %) $N: 286$	853 (92 %) $N: 925$
Nedovoljno voća i povrća	258 (59 %) $N: 435$			884 (91 %) $N: 976$
Pasivno pušenje	275 (96 %) $N: 286$	53 (50 %) $N: 107$		
Konzumacija alkohola				45 (5 %) $N: 954$
Nezadovoljstvo zdravljem		14 (8 %) $N: 165$		134 (14 %) $N: 980$
Neuroticizam	112 (28 %) $N: 394$	55 (34 %) $N: 164$		
Simptomi depresije	1 (0 %) $N: 424$			51 (5 %) $N: 973$

drugom i četvrtom izvoru više od polovice ispitanika bilo starije od 40 godina (43 u drugom i 50 u četvrtom izvoru). Mlađi su ispitanici imali i niži indeks tjelesne mase (medijan 21 i 22 u prvom i trećem izvoru, u usporedbi s medijanom od 24 i 26 u drugom i četvrtom izvoru).

U svim izvorima prevladavali su ispitanici iz gradskog okružja (67 % - 92 %). U izvorima u kojima je stupanj obrazovanja ispitanika bio poznat, ispitanici s barem višom stručnom spremom bili su u manjini (udjeli od 8 %, 29 % i 49 %).

U kućanstvima s centralnim grijanjem živjela je većina ispitanika u prvom i drugom izvoru (67 % i 85 %), za razliku od četvrtog izvora, gdje je takvih ispitanika bilo 31 %. U prenapučenim uvjetima (kućanstvo s manje od jedne sobe po osobi) živjelo je više od trećine ispitanika (45 % u prvom i 40 % u drugom izvoru).

Alergijske bolesti u roditelja zabilježene su u prva tri izvora. Prevalencije su se kretale od 9 % u izvoru podataka iz najstarijeg istraživanja (treći izvor) do 25 % u izvoru podataka iz najnovijeg istraživanja (prvi izvor). Alergijske bolesti u djetinjstvu zabilježene su samo za treći izvor podataka, gdje su pronađene u 7 % ispitanika.

Podaci o kućnim ljubimcima bili su poznati za prva tri izvora. Najzastupljeniji kućni ljubimci bili su psi (u 24 % - 49 % ispitanika), i mačke (u 23 % - 36 % ispitanika), dok su ptice i glodavci bili dosta rjeđi u sva tri izvora: 5 % - 12 % ispitanika držalo je ptice, a 2 % - 7 % glodavce kao kućne ljubimce. Držanje domaćih životinja (perad, svinje, stoka) bilo je u pravilu vezano za ruralna okruženja te su tako ove životinje bile zastupljenije u izvorima s više ispitanika sa sela: 11 % i 13 % za ispitanike u prvom i trećem izvoru te samo 3 % za ispitanike u drugom izvoru podataka.

Prevalencija pušača kretala se od 27 % u četvrtom izvoru do 42 % u trećem izvoru. Pasivno su pušili gotovo svi nepušači u prvom izvoru (96 %) te polovica nepušača u drugom izvoru (50 %).

Od ispitanika s poznatim podacima, većina ih je unosila nedovoljno voća i povrća (59 % u prvom izvoru i 91 % u četvrtom). S druge strane, vrlo je malo ispitanika prijavilo prekomjernu konzumaciju alkohola (5 % u četvrtom izvoru podataka). Udio ispitanika koji su zadovoljili kriterije Svjetske zdravstvene organizacije o dovoljnoj tjelesnoj aktivnosti bio je vrlo visok u prvom i četvrtom izvoru (redom 94 % i 92 %), dok je u trećem izvoru onih koji su se redovito bavili ikakvom tjelesnom aktivnošću bilo samo 28 %.

Vrlo je malo ispitanika imalo simptome depresije (samo jedan ispitanik u prvom te 5 % u četvrtom izvoru podataka). Također, malo ih je bilo nezadovoljnih svojim zdravljem (8 % u drugom izvoru i 14 % u četvrtom izvoru).

5.2.2 Evaluacija metoda na testnom skupu

U ovom dijelu disertacije opisuju se rezultati evaluacije tri vrste modela (logistička regresija i Bayesova mreža izgrađeni na pojedinačnim izvorima podataka te Bayesova mreža izgrađena koristeći sva četiri izvora podataka). Modeli su izgrađeni samo na dijelu podataka iz svakog epidemiološkog izvora (skupovi za treniranje), kako je opisano u poglavlju 4.2.2, a na preostalim podacima iz svakog od izvora (skupovi za testiranje) su evaluirani.

Budući da je cilj ovog dijela rada bio samo odabir najpogodnijeg načina analize ovih epidemioloških izvora te da je samo dio podataka korišten za izradu modela, dobiveni modeli su u ovom poglavlju prikazani bez interpretacije. Naglasak je bio na analizi prediktivne točnosti modela, dok su prikaz i interpretacija medicinskog značaja završnih modela (modela dobivenih iz svih podataka) dani u poglavljima 5.2.3 i 6.

Modeli dobiveni logističkom regresijom skupova za treniranje prikazani su u Tablici 11, dok Slike 7 – 10 prikazuju Bayesove mreže dobivene iz svakog izvora pojedinačno. Budući da su svi ispitanici iz skupa za treniranje prvog i trećeg izvora bili u istoj dobnoj kategoriji (mlađi od 45 godina), za procjenu parametara i evaluaciju korišten je model bez dobi. Modeli za prvi izvor nisu uključili niti simptome depresije jer je samo jedan ispitanik imao simptome.

COmbINE metodom je iz mreža prikazanih na Slikama 7 – 10 dobiven model prikazan na Slici 11.

Kao završni model Bayesove mreže iz sva četiri izvora, izveden iz modela sa Slike 11, odabran je model prikazan na Slici 12. Od veza koje su označene kao moguće, ostavljene su samo one koje je bilo moguće procijeniti (one gdje su povezane varijable bile mjerene barem u jednom izvoru podataka zajedno). Jedna veza koja je bila označena kao sigurna (povezanost alergijskih bolesti u djetinjstvu s nezadovoljstvom zdravljem) izostavljena je, također zbog toga što ove dvije varijable nisu bile mjerene

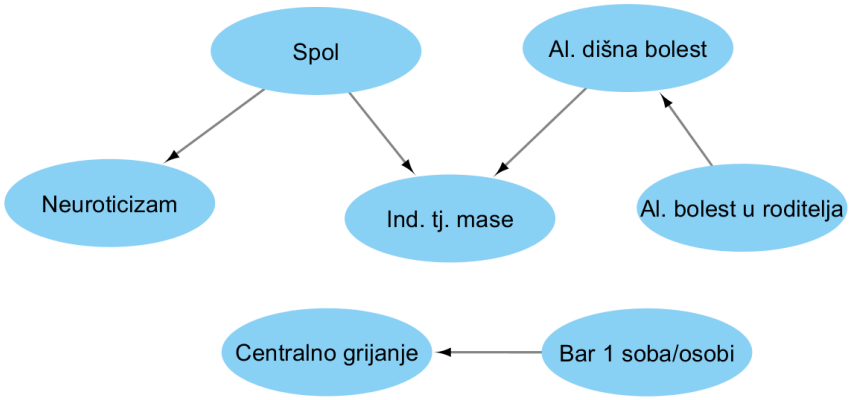
Tablica 11: Omjeri šansi za alergijske dišne bolesti s 95%-tnim intervalima pouzdanosti i p-vrijednostima – višestruke logističke regresije (skupovi za treniranje). Konstantni član prikazan je u izvornom obliku.

Čimbenik	1. izvor	2. izvor	3. izvor
Konstanta	-1,97	-1,55	-3,90
Muški spol	1,85 [0,96 - 3,54] $p:0,064$	1,61 [0,57 - 4,55] $p:0,372$	1,83 [0,51 - 6,61] $p:0,354$
Dob ≥45 godina		1,83 [0,77 - 4,36] $p:0,169$	
Pušenje	1,77 [0,96 - 3,27] $p:0,069$	0,32 [0,12 - 0,86] $p:0,023^*$	0,12 [0,02 - 0,76] $p:0,024^*$
Indeks tjelesne mase >25	0,31 [0,10 - 0,96] $p:0,042^*$	0,64 [0,26 - 1,58] $p:0,331$	0,29 [0,04 - 2,19] $p:0,231$
VSS/VŠS obrazovanje		1,07 [0,47 - 2,46] $p:0,871$	3,81 [1,02 - 14,23] $p:0,047^*$
Život u gradu	0,81 [0,41 - 1,59] $p:0,533$	2,30 [0,42 - 12,49] $p:0,336$	1,08 [0,19 - 6,33] $p:0,928$
Barem 1 soba po osobi	1,19 [0,66 - 2,17] $p:0,564$	1,37 [0,58 - 3,26] $p:0,471$	
Centralno grijanje	1,79 [0,94 - 3,39] $p:0,076$	0,64 [0,19 - 2,14] $p:0,471$	
Al. bolesti u roditelja	3,12 [1,68 - 5,80] $p<0,001^*$	2,50 [0,73 - 8,48] $p:0,143$	5,95 [1,45 - 24,44] $p:0,013^*$
Al. bolesti u djetinjstvu			34,90 [5,38 - 226,30] $p<0,001^*$
Dovoljna tj. aktivnost	0,83 [0,25 - 2,73] $p:0,758$		3,00 [0,78 - 11,54] $p:0,110$
Nedovoljno voća i povrća	1,06 [0,59 - 1,93] $p:0,841$		
Nezadovoljstvo zdravljem		1,42 [0,36 - 5,69] $p:0,617$	
Neuroticizam	0,72 [0,36 - 1,45] $p:0,356$	1,09 [0,45 - 2,65] $p:0,845$	

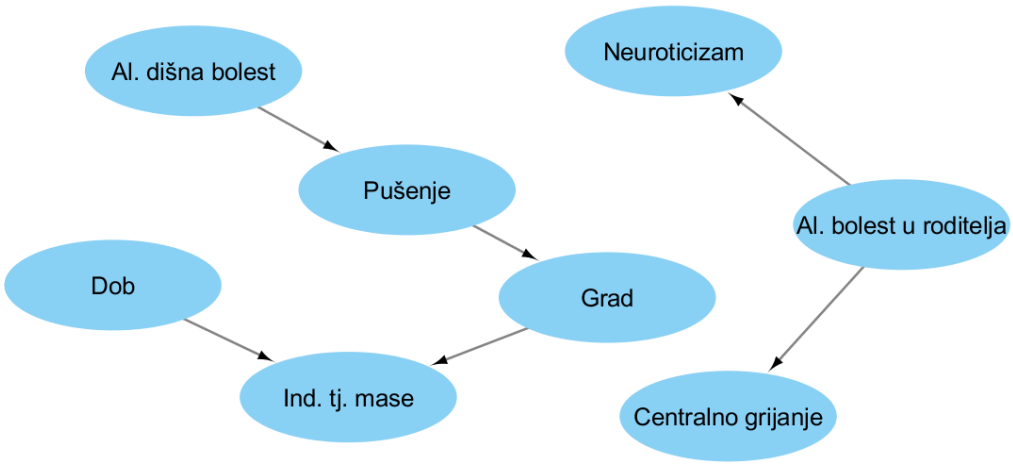
* $p<0,05$

zajedno niti u jednom izvoru. Osim toga, za dvije je veze sa Slike 11 obrnut smjer (povezanost stupnja obrazovanja i pušenja te alergijske dišne bolesti i alergijske bolesti u roditelja) jer su originalni smjerovi bili logički nemogući.

Parametri za alergijske dišne bolesti procijenjeni su G-MA metodom, na isti način kako je opisano u poglavlju 4.2.1.4. Preostali su parametri procijenjeni običnom meta-analizom (model s fiksnim efektima) ako je bilo više izvora u kojima su povezane

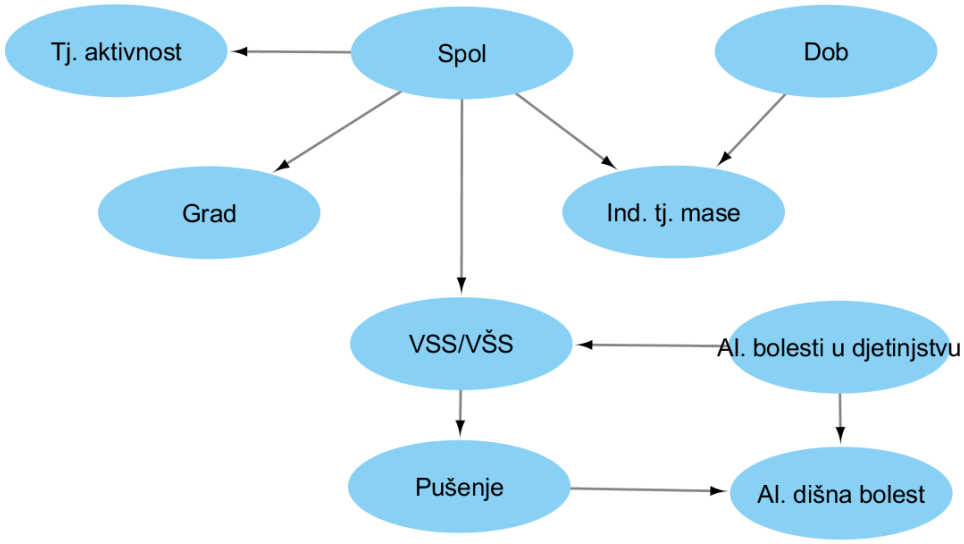


Slika 7: Bayesova mreža za skup za treniranje 1. izvora podataka. Čimbenici koji nisu bili povezani ni s jednom drugom varijablom u mreži nisu prikazani. Slika je napravljena u programu Cytoscape.

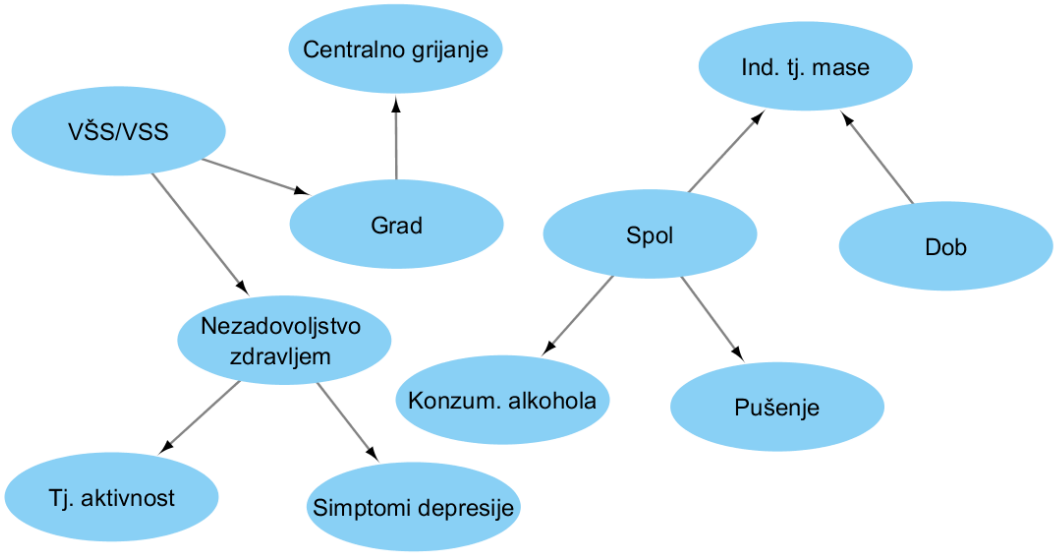


Slika 8: Bayesova mreža za skup za treniranje 2. izvora podataka. Čimbenici koji nisu bili povezani ni s jednom drugom varijablom u mreži nisu prikazani. Slika je napravljena u programu Cytoscape.

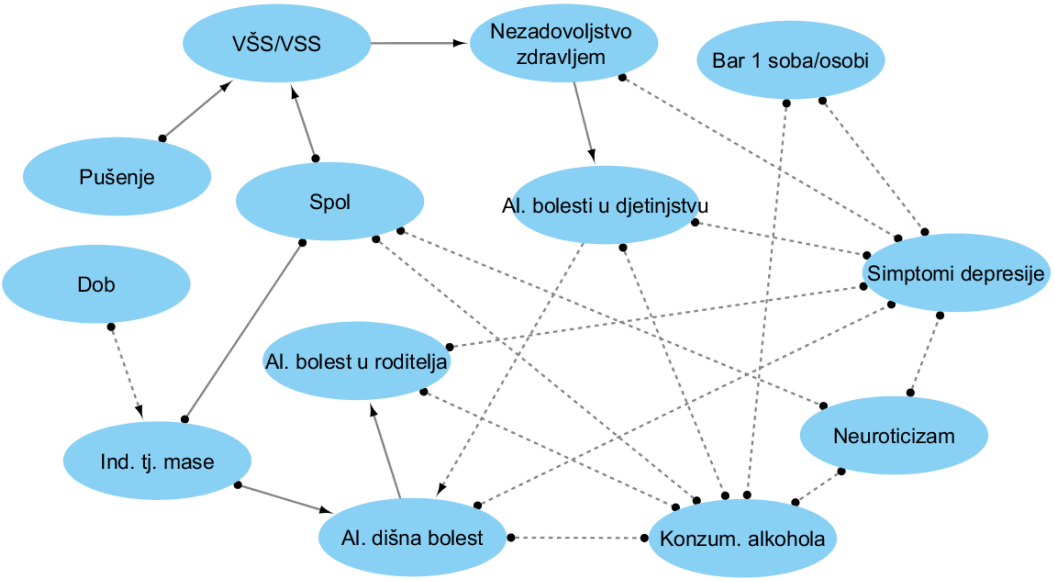
varijable bile dostupne, ili analizom samo jednog izvora u suprotnom. Za indeks tjelesne mase korišten je model logističke regresije s interakcijom između dobi i spola. Razmotrene su i interakcije između alergijskih bolesti u roditelja, alergijskih bole-



Slika 9: Bayesova mreža za skup za treniranje 3. izvora podataka. Čimbenici koji nisu bili povezani ni s jednom drugom varijablom u mreži nisu prikazani. Slika je napravljena u programu Cytoscape.



Slika 10: Bayesova mreža za skup za treniranje 4. izvora podataka. Čimbenici koji nisu bili povezani ni s jednom drugom varijablom u mreži nisu prikazani. Slika je napravljena u programu Cytoscape.

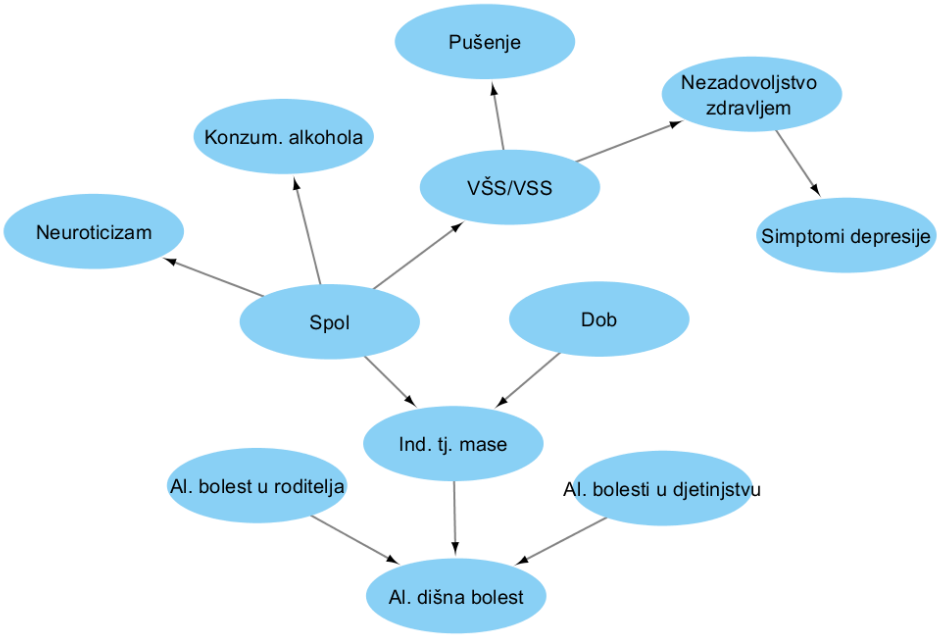


Slika 11: Mreža dobivena primjenom COmbINE metode na mreže sa Slika 7 - 10. Čimbenici koji nisu bili povezani ni s jednom drugom varijablom u mreži nisu prikazani. Veze koje nisu sigurne su iscrtane (--), dok veze s nesigurnim smjerovima počinju ili završavaju točkom (•). Slika je napravljena u programu Cytoscape.

sti u djetinjstvu i prekomjerne tjelesne težine pri utjecaju na prisutnost alergijske dišne bolesti. Međutim, interakcija između alergijskih bolesti u roditelja i prekomjerne tjelesne težine nije bila značajna u prva dva izvora podataka, dok interakcije u trećem izvoru podataka nisu ispitane zbog preniskih prevalencija alergijskih dišnih bolesti, alergijskih bolesti u roditelja i alergijskih bolesti u djetinjstvu. Zbog toga je u procjeni parametara za prisutnost alergijske dišne bolesti korišten model bez interakcija.

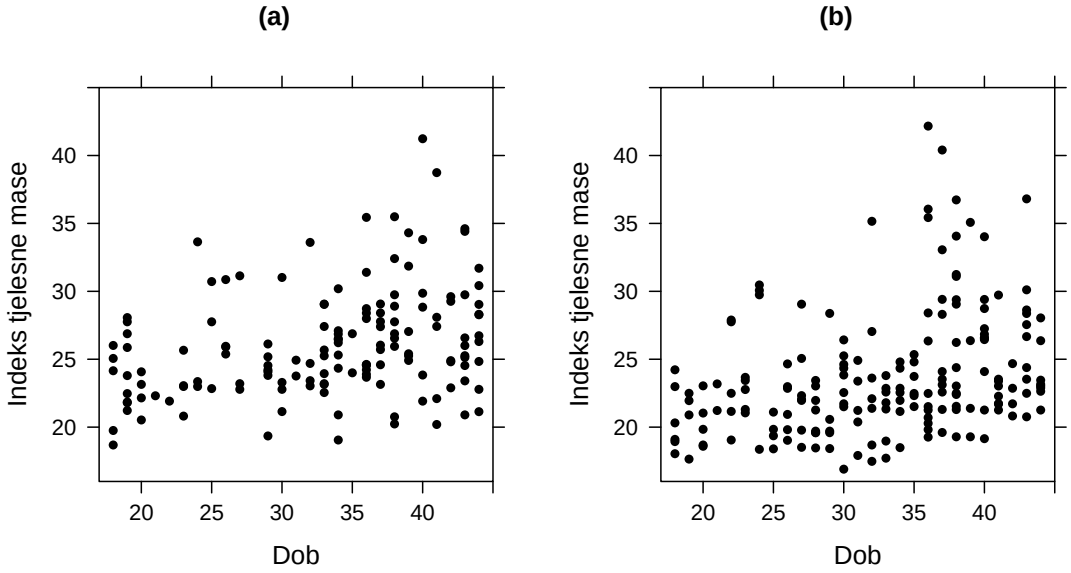
Iako su spol i stupanj obrazovanja bili dostupni u tri izvora, za procjenu njihove povezanosti korišten je samo četvrti izvor. Naime, zbog specifičnosti radnih mjesta ispitanika u drugom i trećem izvoru, bilo je očekivano da će ovi izvori dati pristranu sliku o utjecaju spola na stupanj obrazovanja.

Također, samo su drugi i četvrti izvor korišteni za procjenu vjerojatnosti povišenog indeksa tjelesne mase u odnosu na dob i spol jer su skupovi za treniranje prvog i trećeg izvora podataka sadržavali većinom mlade ispitanike (100 % i 94 % ispitanika



Slika 12: Završna mreža za sva četiri izvora, izvedena iz mreže prikazane na Slici 11. Slika je napravljena u programu Cytoscape.

mlađih od 35 godina) kod kojih nije toliko izražen utjecaj dobi na indeks tjelesne mase kao u starijih ispitanika iz iste dobne skupine, mlađih od 45 godina (Slika 13). Za procjenu spolne i dobne strukture populacije korišten je samo četvrti izvor, također zbog specifične dobne i spolne strukture ostalih izvora podataka.



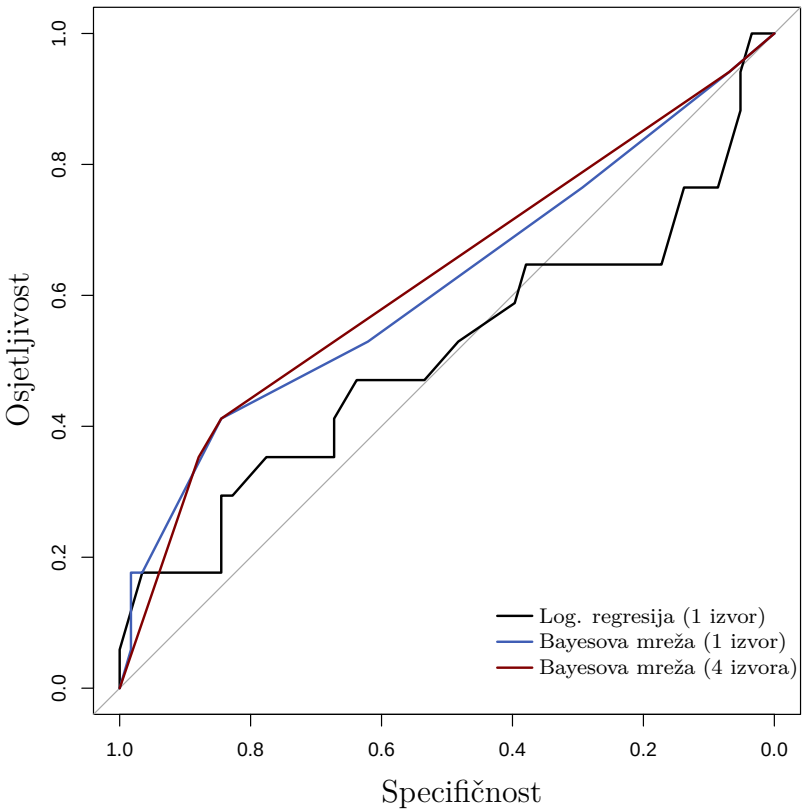
Slika 13: Odnos indeksa tjelesne mase i dobi kod (a) muškaraca i (b) žena u četvrtom izvoru podataka.

Rezultati evaluacije: 1. izvor

Skup za testiranje tri odabrana modela za prvi izvor podataka sastojao se od 88 ispitanika, od kojih je 75 (17 s alergijskom bolesti i 58 bez) s potpunim podacima uključeno u daljnju evaluaciju. Rezultati evaluacije nalaze se na Slici 14 i u Tablici 12. ROC krivulje za modele Bayesovih mreža ponašale su se vrlo slično: površina ispod krivulje bila je 0,62 za model jednog izvora podataka te 0,63 za model sa četiri izvora podataka. Model logističke regresije imao je primjetno manju površinu ispod ROC krivulje (0,51). Maksimalne točnosti modela nisu se jako razlikovale: kretale su se od 0,77 za model Bayesove mreže četiri izvora do 0,80 za Bayesovu mrežu jednog izvora. Maksimalna točnost za model logističke regresije postignuta je kad je uzeta granična vrijednost vjerojatnosti bolesti (vrijednost takva da su svi s vjerojatnosti manjom od granične klasificirani kao zdravi) u rasponu od 0,48 do 0,53. Za ove granične vrijednosti pozitivne prediktivne vrijednosti kretale su se od 0,60 do 1, a negativne od 0,78 do 0,80 (Tablica 12). Maksimalna točnost Bayesove mreže jednog izvora postignuta je za graničnu vrijednost vjerojatnosti od 0,40 te su pozitivna i negativna prediktivna vrijednost u ovom slučaju bile 0,75 i 0,80. Model četiri izvora je imao najveću postignutu točnost 0,77 za granične vrijednosti vjerojatnosti od 0,32 i više. U tom su slučaju svi ispitanici klasificirani kao zdravi pa pozitivnu prediktivnu vrijednost nije bilo moguće izračunati, a negativna prediktivna vrijednost bila je jednaka točnosti (0,77).

Rezultati evaluacije: 2. izvor

Skup za testiranje modela za drugi izvor podataka sastojao se od 33 ispitanika (7 s alergijskom bolesti i 26 bez) koji su svi imali potrebne podatke za daljnju evaluaciju. Rezultati evaluacije prikazani su na Slici 15 i u Tablici 13. Najbolji



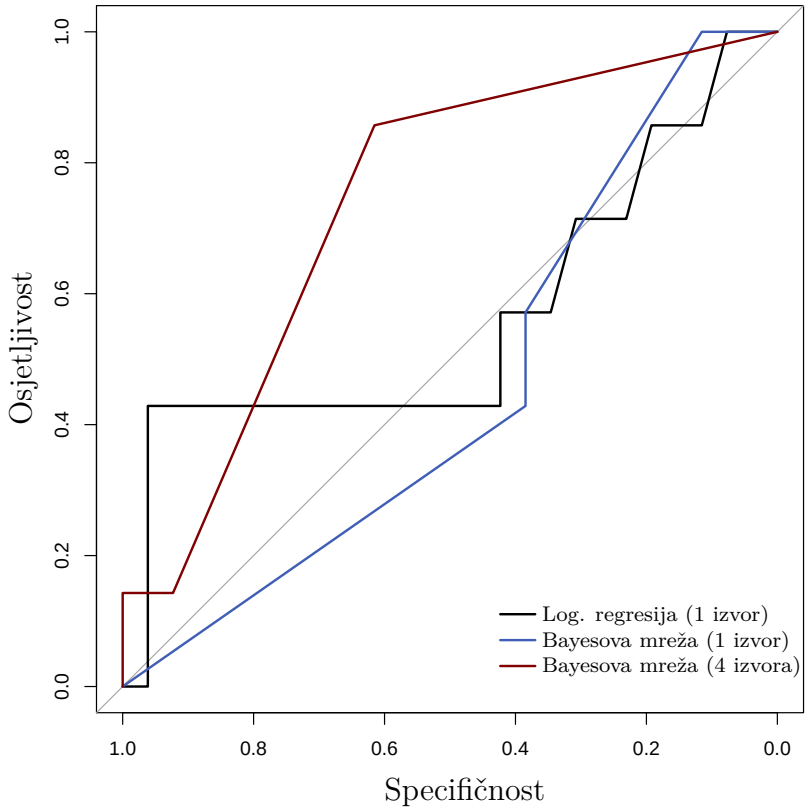
Slika 14: Usporedba osjetljivosti i specifičnosti (ROC analiza) tri odabrane metode na skupu za testiranje prvog izvora podataka.

Tablica 12: Evaluacija tri metode na skupu za testiranje prvog izvora podataka. PPV: pozitivna prediktivna vrijednost, NPV: negativna prediktivna vrijednost.

Metoda	Površina ispod ROC krivulje	Maksimalna točnost	PPV	NPV
Logistička regresija (1 izvor)	$0,51$	$0,79$	$0,60$	$0,80$
			$0,67$	$0,79$
			1	$0,78$
Bayesova mreža (1 izvor)	$0,62$	$0,80$	$0,75$	$0,80$
Bayesova mreža (4 izvora)	$0,63$	$0,77$	–	$0,77$

omjer osjetljivosti i specifičnosti pokazao je model Bayesove mreže četiri izvora (površina ispod ROC krivulje 0,74), dok su druga dva modela imala značajno nižu površinu ispod krivulje (0,56 za logističku regresiju i 0,46 za Bayesovu mrežu jednog izvora). S druge strane, logistička regresija je imala najveću postignutu točnost:

0,85, postignuta u slučajevima kad je kao vjerojatnost bolesti koja odjeljuje bolesne od zdravih uzeta vrijednost u rasponu od 0,45 i 0,48. U ovim je slučajevima pozitivna prediktivna vrijednost bila 0,75, a negativna 0,86. Druga najbolja metoda po točnosti bila je Bayesova mreža četiri izvora. Najbolja postignuta točnost ove metode bila je 0,82 za graničnu vrijednost vjerojatnosti 0,31 i u tom slučaju je pozitivna prediktivna vrijednost bila 1, a negativna 0,81. Model Bayesove mreže jednog izvora je imao najveću postignutu točnost 0,79 kad je granična vrijednost vjerojatnosti bila 0,37 i više. U tom su slučaju svi ispitanici klasificirani kao zdravi pa pozitivnu prediktivnu vrijednost nije bilo moguće izračunati, a negativna prediktivna vrijednost bila je 0,79.



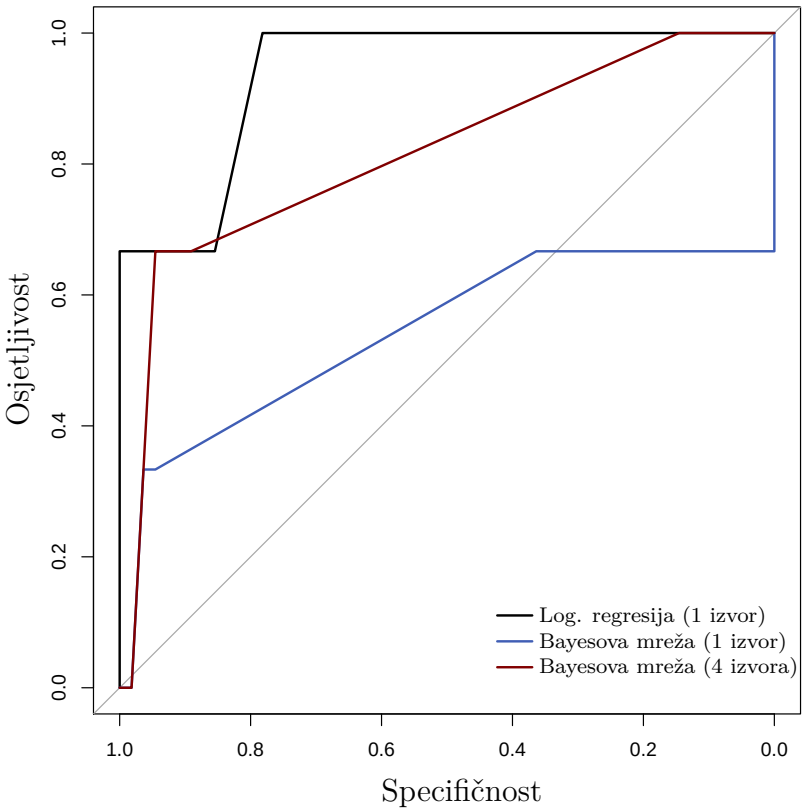
Slika 15: Usporedba osjetljivosti i specifičnosti (ROC analiza) tri odabrane metode na skupu za testiranje drugog izvora podataka.

Tablica 13: Evaluacija tri metode na skupu za testiranje drugog izvora podataka. PPV: pozitivna prediktivna vrijednost, NPV: negativna prediktivna vrijednost.

Metoda	Površina ispod ROC krivulje	Maksimalna točnost	PPV	NPV
Logistička regresija (1 izvor)	0,56	0,85	0,75	0,86
Bayesova mreža (1 izvor)	0,46	0,79	–	0,79
Bayesova mreža (4 izvora)	0,74	0,82	1	0,81

Rezultati evaluacije: 3. izvor

Skup za testiranje modela za treći izvor podataka sastojao se od 62 ispitanika, od kojih je 58 (3 s alergijskom bolesti i 55 bez) imalo sve potrebne podatke za daljnju evaluaciju. Rezultati evaluacije prikazani su na Slici 16 i u Tablici 14. U ovom se slučaju najboljim po svim parametrima pokazao model logističke regresije, s površinom ispod ROC krivulje od 0,94, najvećom postignutom točnosti od 0,98 (za vrijednosti vjerojatnosti u rasponu od 0,70 do 0,74) te pozitivnom prediktivnom vrijednosti od 1 i negativnom od 0,98 za tu točnost. Drugi najbolji model bio je model Bayesove mreže četiri izvora (površina ispod ROC krivulje 0,82), dok je model Bayesove mreže jednog izvora imao najgori omjer osjetljivosti i specifičnosti (površina ispod ROC krivulje 0,54). Oba modela Bayesove mreže najveću su točnost postigla kad su sve ispitanike klasificirale kao zdrave (granična vjerojatnost od 0,81 i više za model jednog izvora te 0,77 i više za model četiri izvora). U tom su slučaju i točnost i negativna prediktivna vrijednost za oba modela iznosila 0,95, dok pozitivnu prediktivnu vrijednost nije bilo moguće izračunati.



Slika 16: Usporedba osjetljivosti i specifičnosti (ROC analiza) tri odabrane metode na skupu za testiranje trećeg izvora podataka.

Tablica 14: Evaluacija tri metode na skupu za testiranje trećeg izvora podataka. PPV: pozitivna prediktivna vrijednost, NPV: negativna prediktivna vrijednost.

Metoda	Površina ispod ROC krivulje	Maksimalna točnost	PPV	NPV
Logistička regresija (1 izvor)	$0,94$	$0,98$	1	$0,98$
Bayesova mreža (1 izvor)	$0,54$	$0,95$	–	$0,95$
Bayesova mreža (4 izvora)	$0,82$	$0,95$	–	$0,95$

5.2.3 Rizični čimbenici za alergijske dišne bolesti

5.2.3.1 Rizični čimbenici za alergijske dišne bolesti: 1. izvor podataka

Rezultati jednostrukih logističkih regresija za sva tri izvora prikazani su u Tablici 15. Iako su simptomi depresije bili zabilježeni u prvom izvoru, gdje su ispitanici bili studenti preddiplomskih studija, nisu uključeni u daljnje analize jer je samo jedan ispitanik imao simptome (Tablica 10). Jedini čimbenik koji je pokazao statistički značajnu povezanost s alergijskim dišnim bolestima u prvom izvoru bile su alergijske bolesti u roditelja: ispitanici kojima su roditelji imali neku od alergijskih bolesti imali su 2,86 puta veće izglede (omjer šansi 2,86 s 95 %-tnim intervalom pouzdanosti [1,76 – 4,65]) za alergijsku dišnu bolest nego ostali ispitanici.

Rezultati modela višestruke logističke regresije prikazani su u Tablici 16. Osim simptoma depresije, u model nije bila uključena ni dob, jer se većina ispitanika nalazila u uskom dobnom rasponu (interkvartilni raspon 18–19 godina), te neuroticizam jer za 10 % ispitanika nisu bile poznate vrijednosti ove varijable (Tablica 10). Nijedan od ova dva čimbenika nije pokazao značajnu povezanost s alergijskim dišnim bolestima u jednostrukim modelima.

Alergijske bolesti u roditelja ostale su najsnažniji prediktor alergijskih dišnih bolesti i u modelu višestruke logističke regresije s omjerom šansi od 3,06 (95 %-tni interval pouzdanosti [1,84 – 5,11]). Značajno veće izglede za alergijske dišne bolesti imali su i muški ispitanici (omjer šansi 2,09 [1,16 – 3,78]), dok je povećanje indeksa tjelesne mase smanjivalo šanse za alergijske dišne bolesti (omjer šansi 0,88 [0,79 – 0,97]). Ostali čimbenici nisu pokazali značajnu povezanost s alergijskim dišnim bolestima ni u modelu jednostruke niti višestruke regresije.

Analizom Bayesove mreže dobivene su iste povezanosti kao na skupu za treniranje (Slika 17). U analizi su alergijske dišne bolesti bile izravno povezane samo s alergijskim bolestima u roditelja, koje su izdvojene kao uzrok alergijskih dišnih

Tablica 15: Omjeri šansi za alergijske dišne bolesti s 95%-tnim intervalima pouzdanosti i p-vrijednostima za svaki od čimbenika iz epidemioloških izvora podataka – jednostruke logističke regresije

Čimbenik	1. izvor	2. izvor	3. izvor
Muški spol	1,39 [0,85 – 2,26] $p: 0,189$	1,31 [0,58 – 2,97] $p: 0,515$	1,18 [0,48 – 2,90] $p: 0,713$
Dob	1,01 [0,88 – 1,17] $p: 0,881$	1,01 [0,97 – 1,05] $p: 0,734$	1,02 [0,94 – 1,10] $p: 0,689$
Pušenje	0,95 [0,59 – 1,53] $p: 0,828$	0,50 [0,23 – 1,08] $p: 0,077$	0,31 [0,10 – 0,94] $p: 0,039^*$
Indeks tjelesne mase	0,92 [0,84 – 1,00] $p: 0,057$	0,93 [0,84 – 1,02] $p: 0,125$	0,93 [0,80 – 1,09] $p: 0,372$
VSS/VŠS obrazovanje		1,34 [0,68 – 2,65] $p: 0,402$	2,37 [0,97 – 5,80] $p: 0,059$
Život u gradu	1,33 [0,77 – 2,31] $p: 0,305$	2,47 [0,53 – 11,48] $p: 0,250$	2,51 [0,57 – 11,10] $p: 0,224$
Veličina kućanstva	1,13 [0,75 – 1,69] $p: 0,557$	1,11 [0,63 – 1,95] $p: 0,712$	
Centralno grijanje	1,52 [0,92 – 2,53] $p: 0,105$	0,74 [0,29 – 1,86] $p: 0,520$	
Al. bolesti u roditelja	2,86 [1,76 – 4,65] $p < 0,001^*$	2,03 [0,76 – 5,42] $p: 0,159$	6,84 [2,47 – 18,97] $p < 0,001^*$
Al. bolesti u djetinjstvu			13,28 [4,63 – 38,10] $p < 0,001^*$
Dovoljna tj. aktivnost	1,21 [0,45 – 3,31] $p: 0,705$		2,21 [0,88 – 5,56] $p: 0,091$
Nedovoljno voća i povrća	0,91 [0,58 – 1,45] $p: 0,704$		
Nezadovoljstvo zdravljem		1,49 [0,47 – 4,71] $p: 0,497$	
Neuroticizam	0,91 [0,53 – 1,56] $p: 0,741$	1,23 [0,61 – 2,52] $p: 0,563$	

* $p < 0,05$

bolesti, te s indeksom tjelesne mase, pri čemu je promjena u indeksu tjelesne mase kod ispitanika s alergijskom dišnom bolesti označen kao posljedica bolesti. Spol je u mreži izdvojen samo kao drugi čimbenik koji utječe na indeks tjelesne mase, dok izravna povezanost s alergijskim dišnim bolestima nije dobivena.

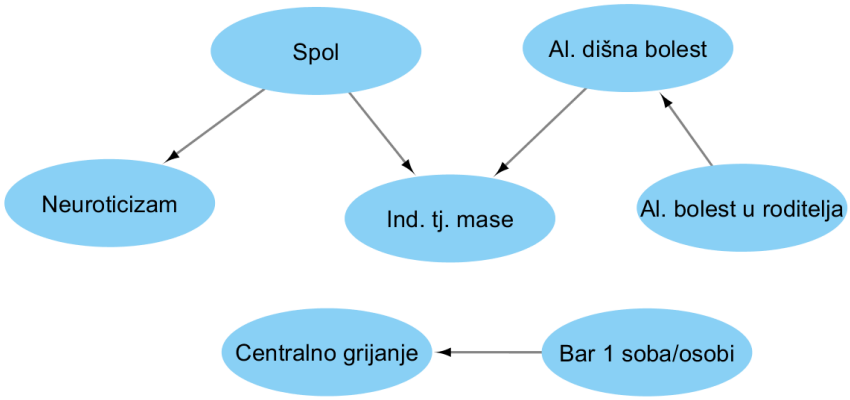
Procijenjene vjerojatnosti za ovaj model dane su u Tablici 17. Alergijske bolesti u roditelja povećavale su vjerojatnost prisutnosti alergijske dišne bolesti: procijenjena

Tablica 16: Omjeri šansi za alergijske dišne bolesti s 95%-tnim intervalima pouzdanosti i p-vrijednostima – višestruke logističke regresije. Konstantni član prikazan je u izvornom obliku.

Čimbenik	1. izvor	2. izvor	3. izvor
Konstanta	0,67	-0,26	-1,95
Muški spol	2,09 [1,16 – 3,78] $p: 0,014^*$	1,93 [0,76 – 4,88] $p: 0,166$	1,84 [0,50 – 6,83] $p: 0,359$
Dob		1,02 [0,98 – 1,07] $p: 0,364$	1,02 [0,89 – 1,17] $p: 0,748$
Pušenje	0,90 [0,53 – 1,53] $p: 0,699$	0,50 [0,22 – 1,12] $p: 0,092$	0,16 [0,03 – 0,75] $p: 0,021^*$
Indeks tjelesne mase	0,88 [0,79 – 0,97] $p: 0,014^*$	0,91 [0,81 – 1,02] $p: 0,118$	0,86 [0,67 – 1,09] $p: 0,211$
VSS/VŠS obrazovanje		0,92 [0,43 – 1,96] $p: 0,823$	3,32 [0,97 – 11,30] $p: 0,055$
Život u gradu	1,09 [0,61 – 1,96] $p: 0,766$	2,73 [0,55 – 13,53] $p: 0,219$	2,03 [0,34 – 12,05] $p: 0,436$
Barem 1 soba po osobi	1,30 [0,78 – 2,16] $p: 0,307$	1,20 [0,56 – 2,58] $p: 0,645$	
Centralno grijanje	1,35 [0,78 – 2,33] $p: 0,285$	0,63 [0,23 – 1,75] $p: 0,380$	
Al. bolesti u roditelja	3,06 [1,84 – 5,11] $p < 0,001^*$	1,97 [0,68 – 5,69] $p: 0,211$	5,83 [1,47 – 23,13] $p: 0,012^*$
Al. bolesti u djetinjstvu			28,15 [6,59 – 120,31] $p < 0,001^*$
Dovoljna tj. aktivnost	1,07 [0,38 – 3,05] $p: 0,898$		3,81 [1,09 – 13,25] $p: 0,036^*$
Nedovoljno voća i povrća	0,79 [0,48 – 1,29] $p: 0,345$		
Nezadovoljstvo zdravljem		1,43 [0,42 – 4,80] $p: 0,564$	
Neuroticizam		1,22 [0,57 – 2,63] $p: 0,611$	

* $p < 0,05$

vjerojatnost prisutnosti bolesti kod osobe kojoj roditelji imaju alergijsku dišnu bolest bila je 0,38, u usporedbi s vjerojatnosti bolesti od 0,17 kod osobe kojoj roditelji nemaju alergijsku dišnu bolest. S druge strane, prisutnost alergijske dišne bolesti smanjivala je vjerojatnost povišenog indeksa tjelesne mase. Tako je kod muškaraca s alergijskom dišnom bolesti procijenjena vjerojatnost za povišeni indeks tjelesne mase bila 0,11, dok je kod muškaraca bez alergijske dišne bolesti bila 0,20. Vjerojatnost



Slika 17: Bayesova mreža za 1. izvor podataka. Čimbenici koji nisu bili povezani ni s jednom drugom varijablom u mreži nisu prikazani. Slika je napravljena u programu Cytoscape.

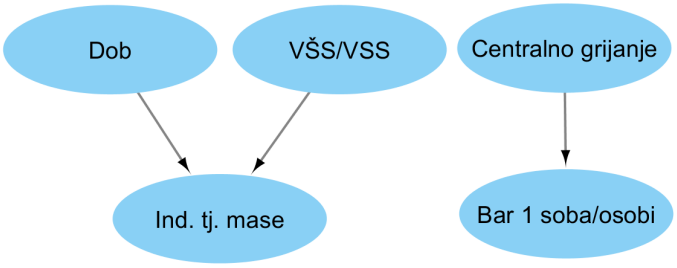
da žena s alergijskom dišnom bolesti ima povišen indeks tjelesne mase bila je 0,04, a ista vjerojatnost za žene bez alergijske dišne bolesti bila je 0,11.

Tablica 17: Procijenjeni parametri za Bayesovu mrežu sa Slike 17.

Čvor	Pridružena vjerojatnost
Spol	$P(\text{spol} = \text{M}) = 0,26$
Al. bolesti u roditelja	$P(\text{al. bolesti u roditelja} = \text{'da'}) = 0,24$
Bar 1 soba po osobi	$P(\text{bar 1 soba po osobi} = \text{'da'}) = 0,55$
Centralno grijanje	$P(\text{centralno grijanje} = \text{'da'} \mid \text{bar 1 soba po osobi} = \text{'da'}) = 0,72$ $P(\text{centralno grijanje} = \text{'da'} \mid \text{bar 1 soba po osobi} = \text{'ne'}) = 0,55$
Neuroticizam	$P(\text{neuroticizam} = \text{'da'} \mid \text{spol} = \text{M}) = 0,13$ $P(\text{neuroticizam} = \text{'da'} \mid \text{spol} = \text{Ž}) = 0,33$
Al. dišna bolest	$P(\text{al. dišna bolest} = \text{'da'} \mid \text{al. bolesti u roditelja} = \text{'da'}) = 0,38$ $P(\text{al. dišna bolest} = \text{'da'} \mid \text{al. bolesti u roditelja} = \text{'ne'}) = 0,17$
Indeks tj. mase (ITM)	$P(\text{ITM} > 25 \mid \text{al. dišna bolest} = \text{'da'},\ \text{spol} = \text{M}) = 0,11$ $P(\text{ITM} > 25 \mid \text{al. dišna bolest} = \text{'da'},\ \text{spol} = \text{Ž}) = 0,04$ $P(\text{ITM} > 25 \mid \text{al. dišna bolest} = \text{'ne'},\ \text{spol} = \text{M}) = 0,20$ $P(\text{ITM} > 25 \mid \text{al. dišna bolest} = \text{'ne'},\ \text{spol} = \text{Ž}) = 0,11$

5.2.3.2 Rizični čimbenici za alergijske dišne bolesti: 2. izvor podataka

Nijedan od čimbenika zabilježenih u drugom izvoru podataka (ispitanici: administrativni radnici) nije pokazao značajnu povezanost s alergijskim dišnim bolestima u ovom izvoru (Tablice 15 i 16). Bayesova mreža za ovaj izvor podataka prikazana je na Slici 18, a procijenjene vjerojatnosti prikazane su u Tablici 18. Analiza je potvrdila rezultate logističke regresije jer ni u mreži nijedan čimbenik nije bio povezan s pojavom alergijskih dišnih bolesti.



Slika 18: Bayesova mreža za 2. izvor podataka. Čimbenici koji nisu bili povezani ni s jednom drugom varijablom u mreži nisu prikazani. Slika je napravljena u programu Cytoscape.

Tablica 18: Procijenjeni parametri za Bayesovu mrežu sa Slike 18.

Čvor	Pridružena vjerojatnost
Dob	$P(\text{dob} \geq 45) = 0,46$
VŠS/VSS	$P(\text{VŠS/VSS} = \text{'da'}) = 0,49$
Centralno grijanje	$P(\text{centralno grijanje} = \text{'da'}) = 0,85$
Bar 1 soba po osobi	$P(\text{bar 1 soba po osobi} = \text{'da'} \mid \text{centralno grijanje} = \text{'da'}) = 0,63$ $P(\text{bar 1 soba po osobi} = \text{'da'} \mid \text{centralno grijanje} = \text{'ne'}) = 0,38$
Indeks tj. mase (ITM)	$P(\text{ITM} > 25 \mid \text{dob} \geq 45,\ \text{VŠS/VSS} = \text{'da'}) = 0,47$ $P(\text{ITM} > 25 \mid \text{dob} \geq 45,\ \text{VŠS/VSS} = \text{'ne'}) = 0,73$ $P(\text{ITM} > 25 \mid \text{dob} < 45,\ \text{VŠS/VSS} = \text{'da'}) = 0,22$ $P(\text{ITM} > 25 \mid \text{dob} < 45,\ \text{VŠS/VSS} = \text{'ne'}) = 0,37$

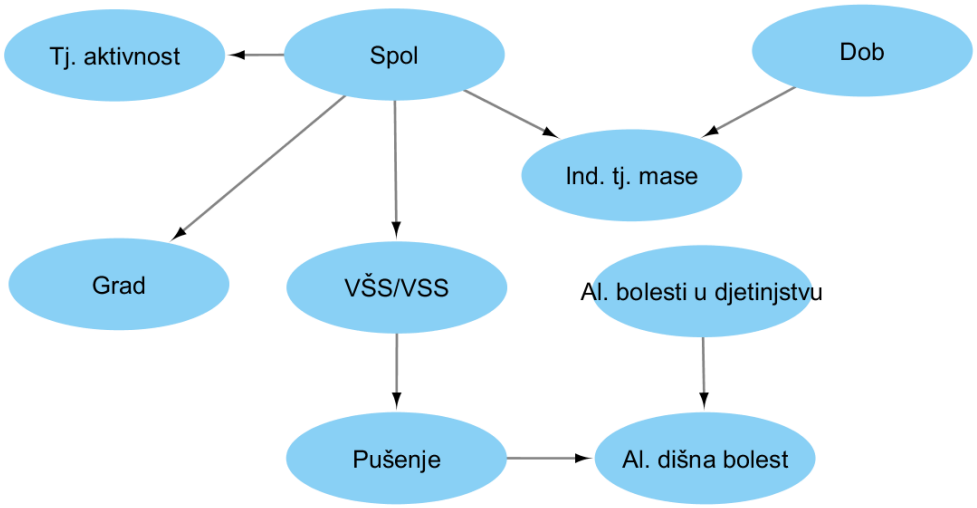
5.2.3.3 Rizični čimbenici za alergijske dišne bolesti: 3. izvor podataka

Treći izvor podataka sačinjavali su podaci iz registra prethodnih pregleda u farmaceutskoj industriji. U analizi jednostrukih povezanosti čimbenika s alergijskim dišnim bolestima (Tablica 15), kao najznačajniji čimbenici pokazali su se alergijske bolesti u roditelja te alergijske bolesti u djetinjstvu. Ispitanici koji su u djetinjstvu imali alergijske bolesti imali su 13,28 puta veće izglede za alergijske dišne bolesti nego ostali ispitanici (omjer šansi 13,28 s 95 %-tnim intervalom pouzdanosti [4,63 – 38,10]). Veće izglede za pojavu alergijskih dišnih bolesti imali su i ispitanici kojima su roditelji imali neku od alergijskih bolesti (omjer šansi 6,84 [2,47 – 18,97]). Pušenje je pak bilo povezano s alergijskim dišnim bolestima u obrnutom smjeru: pušači su imali manje izglede za alergijske dišne bolesti nego nepušači (omjer šansi 0,31 [0,10 – 0,94]).

Ova tri čimbenika potvrđena su i u analizi višestrukom logističkom regresijom (Tablica 16). Veće izglede za alergijske dišne bolesti imali su ispitanici s alergijskim bolestima u djetinjstvu (omjer šansi 28,15 [6,59 – 120,31]) te ispitanici kojima su roditelji imali alergijske bolesti (omjer šansi 5,83 [1,47 – 23,13]). Pušači su i u ovom modelu imali manje izglede za alergijske dišne bolesti nego nepušači (omjer šansi 0,16 [0,03 – 0,75]). Osim ova tri čimbenika, u modelu višestruke regresije značajan čimbenik bila je i tjelesna aktivnost: oni koji su se redovito bavili tjelesnom aktivnošću imali su i veće izglede za alergijske dišne bolesti (omjer šansi 3,81 [1,09 – 13,25]).

U analizi Bayesovom mrežom (Slika 19) kao jedini čimbenici izravno povezani s alergijskim dišnim bolestima izdvojene su alergijske bolesti u djetinjstvu i pušenje. Alergijske bolesti u roditelja nisu bile povezane ni s jednim čimbenikom u mreži, dok je tjelesna aktivnost izravno bila povezana samo sa spolom.

Procijenjeni parametri za ovaj model prikazani su u Tablici 19. Od čimbenika koji su izravno bili povezani s bolesti, alergijske bolesti u djetinjstvu povećavale su



Slika 19: Bayesova mreža za 3. izvor podataka. Čimbenici koji nisu bili povezani ni s jednom drugom varijablom u mreži nisu prikazani. Slika je napravljena u programu Cytoscape.

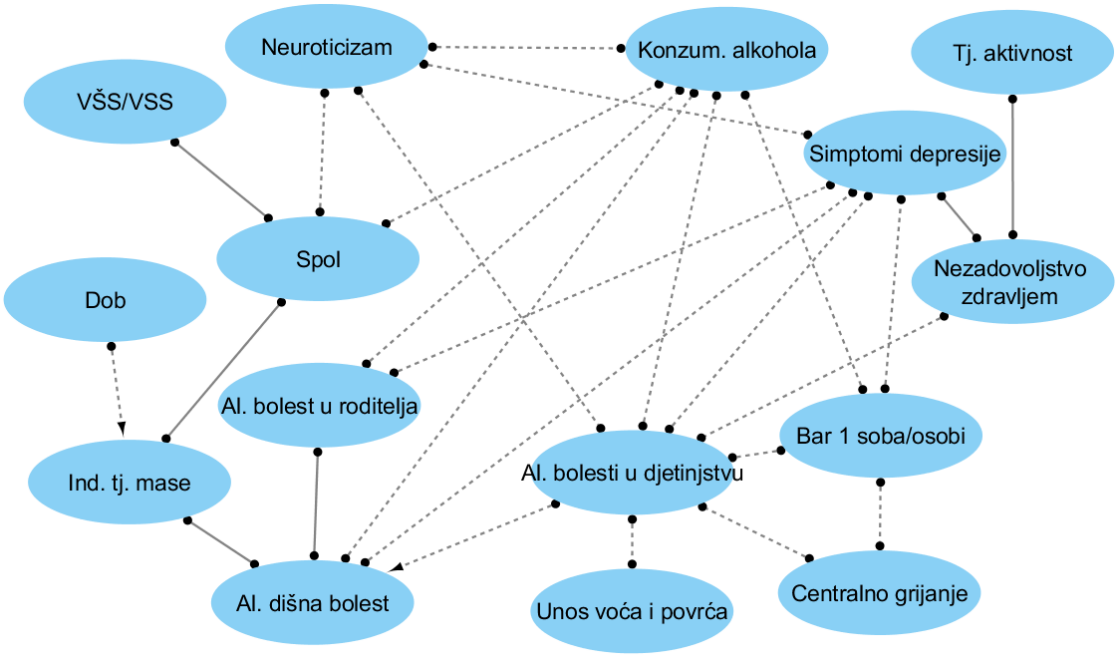
vjerojatnost alergijske dišne bolesti, dok su pušači imali manju vjerojatnost bolesti.

Tablica 19: Procijenjeni parametri za Bayesovu mrežu sa Slike 19. Uvjetne vjerojatnosti za povišeni indeks tjelesne mase za dob od 45 godina i više nije bilo moguće procijeniti jer nije bilo ispitanika te dobi.

Čvor	Pridružena vjerojatnost
Spol	$P(\text{spol} = \text{M}) = 0,53$
Al. bolesti u djetinjstvu	$P(\text{al. bolesti u djetinjstvu} = \text{'da'}) = 0,07$
Pušenje	$P(\text{pušač} = \text{'da'} \mid \text{VŠS/VSS} = \text{'da'}) = 0,26$ $P(\text{pušač} = \text{'da'} \mid \text{VŠS/VSS} = \text{'ne'}) = 0,49$
Indeks tj. mase (ITM)	$P(\text{ITM} > 25 \mid \text{dob} < 45,\ \text{spol} = \text{M}) = 0,34$ $P(\text{ITM} > 25 \mid \text{dob} < 45,\ \text{spol} = \text{Ž}) = 0,08$
Grad	$P(\text{mjesto stanovanja} = \text{'grad'} \mid \text{spol} = \text{M}) = 0,73$ $P(\text{mjesto stanovanja} = \text{'grad'} \mid \text{spol} = \text{Ž}) = 0,87$
Tjelesna aktivnost	$P(\text{redovita tj. aktivnost} = \text{'da'} \mid \text{spol} = \text{M}) = 0,41$ $P(\text{redovita tj. aktivnost} = \text{'da'} \mid \text{spol} = \text{Ž}) = 0,14$
VŠS/VSS	$P(\text{VŠS/VSS} = \text{'da'} \mid \text{spol} = \text{M}) = 0,21$ $P(\text{VŠS/VSS} = \text{'da'} \mid \text{spol} = \text{Ž}) = 0,35$
Al. dišna bolest	$P(\text{al. dišna bolest} = \text{'da'} \mid \text{al. bol. u djetinjstvu} = \text{'da'},\ \text{pušač} = \text{'da'}) = 0,18$ $P(\text{al. dišna bolest} = \text{'da'} \mid \text{al. bol. u djetinjstvu} = \text{'da'},\ \text{pušač} = \text{'ne'}) = 0,67$ $P(\text{al. dišna bolest} = \text{'da'} \mid \text{al. bol u djetinjstvu} = \text{'ne'},\ \text{pušač} = \text{'da'}) = 0,01$ $P(\text{al. dišna bolest} = \text{'da'} \mid \text{al. bol u djetinjstvu} = \text{'ne'},\ \text{pušač} = \text{'ne'}) = 0,07$

5.2.3.4 Zajednička analiza sva četiri izvora podataka

Model dobiven zajedničkom analizom sva četiri izvora podataka koristeći COmbINE metodu prikazan je na Slici 20. Čimbenici koji su u modelu bili izravno povezani s alergijskim dišnim bolestima bili su indeks tjelesne mase, alergijske bolesti u roditelja, alergijske bolesti u djetinjstvu, prekomjerna konzumacija alkohola i simptomi depresije.

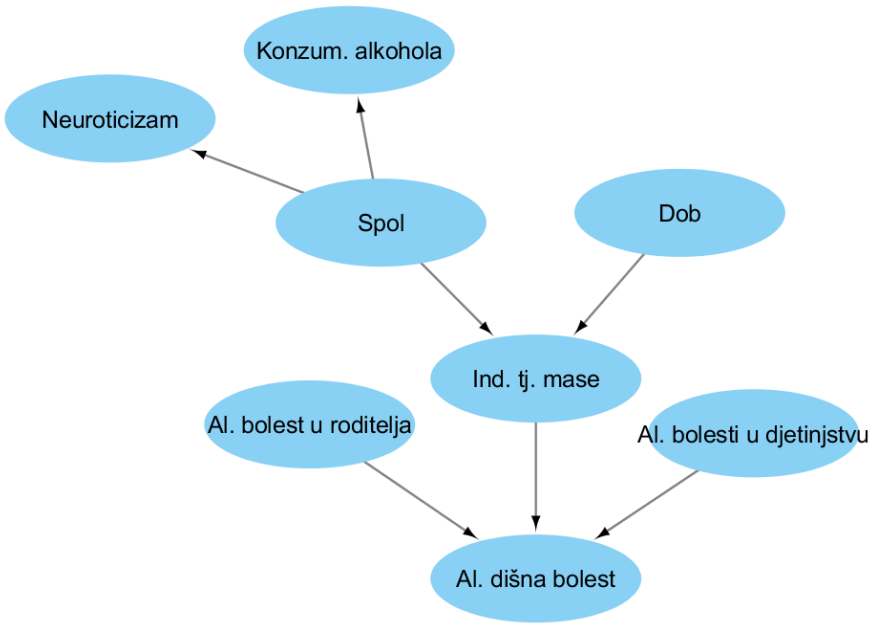


Slika 20: Mreža dobivena primjenom COmbINE metode na mreže iz četiri izvora podataka. Čimbenici koji nisu bili povezani ni s jednom drugom varijablom u mreži nisu prikazani. Veze koje nisu sigurne su iscrtane (--), dok veze s nesigurnim smjerovima počinju ili završavaju točkom (•). Slika je napravljena u programu Cytoscape.

Od ovih pet čimbenika, samo su indeks tjelesne mase i alergijske bolesti u roditelja bili označeni kao sigurno povezani s alergijskim dišnim bolestima, dok su preostala tri čimbenika označena samo kao moguće povezana. Pritom su četiri od ovih pet veza (sve veze osim veze za alergijske bolesti u djetinjstvu) bile označene kao veze kojima nije bilo moguće odrediti smjer (veze oblika •——• ili •·····•). To znači da je moguće više različitih uzročno-posljedičnih interpretacija ovih povezano-

sti. Primjerice, veza 'indeks tjelesne mase •·····• al. dišna bolest' može značiti: 1) alergijska dišna bolest uzrokuje promjenu u indeksu tjelesne mase; 2) indeks tjelesne mase izravno utječe na pojavu alergijske dišne bolesti; ili 3) postoji treći čimbenik (koji nije uključen u postojeće čimbenike) koji uzrokuje i promjenu indeksa tjelesne mase i dovodi do pojave alergijske dišne bolesti.

Kao drugi tip veze, veza 'alergijske bolesti u djetinjstvu •·····▸ alergijske dišne bolesti' znači da je, ako povezanost postoji, moguće da alergijske bolesti u djetinjstvu utječu na pojavu alergijske dišne bolesti kasnije u životu ili da postoji neki nepoznati čimbenik (čimbenik koji se ne nalazi među postojećim čimbenicima u modelu) koji uzrokuje i alergijske bolesti u djetinjstvu i alergijske bolesti kasnije u životu. Za razliku od ostale četiri veze, model ne dopušta interpretaciju da alergijske bolesti kasnije u životu utječu na pojavu alergijskih bolesti u djetinjstvu.



Slika 21: Završna mreža za sva četiri izvora, izvedena iz mreže prikazane na Slici 20. Slika je napravljena u programu Cytoscape.

Model odabran za procjenu parametara prikazan je na Slici 21.

Iz modela sa Slike 20 najprije su uklonjene veze između varijabli koje nisu ni u

jednom izvoru bile mjerene zajedno, zbog čega nije bilo moguće procijeniti jačinu ovih veza. Također su uklonjene veze između simptoma depresije s jedne strane i neuroticizma, alergijskih bolesti u roditelja, prisutnosti alergijske dišne bolesti i veličine kućanstva s druge strane: iako su sve ove varijable bile mjerene u prvom izvoru, zbog vrlo niske prevalencije simptoma depresije u ovom izvoru nije bilo moguće odrediti jačinu niti za ove veze. Sve uklonjene veze bile su označene samo kao moguće.

Nakon uklanjanja ovih veza, u modelu su ostavljene samo one varijable koje su bile povezane s alergijskim dišnim bolestima, bilo izravno (kao što je veza prisutnosti alergijske dišne bolesti i indeksa tjelesne mase), bilo neizravno (putem drugih varijabli, kao što je spol bio povezan s prisutnošću alergijske dišne bolesti putem indeksa tjelesne mase). Primjerice, iako je veza između nezadovoljstva zdravljem i simptoma depresije bila označena kao sigurna te je bilo moguće procijeniti njenu jačinu iz četvrtog izvora, nakon uklanjanja veza koje nije bilo moguće procijeniti niti jedna od ove dvije varijable nije bila povezana s prisutnošću alergijske dišne bolesti, ni izravno niti neizravno, te iz tog razloga ove dvije varijable nisu zadržane u završnom modelu.

Naposljetku, veza između spola i obrazovanja (povezanost označena kao sigurna u modelu sa Slike 20) nije zadržana u završnom modelu jer se u daljnjoj analizi nije pokazala značajnom (omjer šansi za povezanost spola i barem više stručne spreme bio je 1,01 [0,59 – 1,75] u četvrtom izvoru). U analizu povezanosti spola i obrazovanja nisu bili uključeni drugi i treći izvor iz istog razloga kao u poglavlju 5.2.2: zbog specifičnosti radnih mjesta ovih ispitanika te dodatnog isključivanja skupina ispitanika (npr. žena čistačica), očekivana je pristranost u procjeni odnosa spola i obrazovanja iz ovih izvora. Isto tako, kako je detaljnije objašnjeno u poglavlju 5.2.2, za procjenu spolne i dobne strukture populacije korišten je samo četvrti izvor, dok su za procjenu ovisnosti indeksa tjelesne mase korišteni samo drugi i četvrti

izvor podataka.

Dobiveni parametri prikazani su u Tablici 20. Kao što se vidi iz Tablice, alergijske bolesti u roditelja i alergijske bolesti u djetinjstvu povećavale su vjerojatnost prisutnosti alergijske dišne bolesti, dok je povećani indeks tjelesne mase bio povezan s manjom vjerojatnosti prisutnosti bolesti.

Tablica 20: Procijenjeni parametri za Bayesovu mrežu sa Slike 21.

Čvor	Pridružena vjerojatnost
Spol	$P(\text{spol} = \text{M}) = 0,42$
Dob	$P(\text{dob} \geq 45) = 0,63$
Al. bolesti u roditelja	$P(\text{alergijske bolesti u roditelja} = \text{'da'}) = 0,09$
Al. bolesti u djetinjstvu	$P(\text{alergijske bolesti u djetinjstvu} = \text{'da'}) = 0,07$
Konzum. alkohola	$P(\text{pretjerana konzumacija alkohola} = \text{'da'} \mid \text{spol} = \text{M}) = 0,11$ $P(\text{pretjerana konzumacija alkohola} = \text{'da'} \mid \text{spol} = \text{Ž}) = 0,01$
Neuroticizam	$P(\text{neuroticizam} = \text{'da'} \mid \text{spol} = \text{M}) = 0,16$ $P(\text{neuroticizam} = \text{'da'} \mid \text{spol} = \text{Ž}) = 0,35$
Indeks tj. mase (ITM)	$P(\text{ITM} > 25 \mid \text{dob} < 45,\ \text{spol} = \text{M}) = 0,53$ $P(\text{ITM} > 25 \mid \text{dob} < 45,\ \text{spol} = \text{Ž}) = 0,25$ $P(\text{ITM} > 25 \mid \text{dob} \geq 45,\ \text{spol} = \text{M}) = 0,66$ $P(\text{ITM} > 25 \mid \text{dob} \geq 45,\ \text{spol} = \text{Ž}) = 0,62$
Al. dišna bolest	$P(\text{al. dišna bol.} = \text{'da'} \mid \text{al. bol. u rod.} = \text{'da'},\ \text{al. bol. u dj.} = \text{'da'},\ \text{ITM} > 25) = 0,69$ $P(\text{al. dišna bol.} = \text{'da'} \mid \text{al. bol. u rod.} = \text{'da'},\ \text{al. bol. u dj.} = \text{'da'},\ \text{ITM} \leq 25) = 0,81$ $P(\text{al. dišna bol.} = \text{'da'} \mid \text{al. bol. u rod.} = \text{'da'},\ \text{al. bol. u dj.} = \text{'ne'},\ \text{ITM} > 25) = 0,15$ $P(\text{al. dišna bol.} = \text{'da'} \mid \text{al. bol. u rod.} = \text{'da'},\ \text{al. bol. u dj.} = \text{'ne'},\ \text{ITM} \leq 25) = 0,25$ $P(\text{al. dišna bol.} = \text{'da'} \mid \text{al. bol. u rod.} = \text{'ne'},\ \text{al. bol. u dj.} = \text{'da'},\ \text{ITM} > 25) = 0,47$ $P(\text{al. dišna bol.} = \text{'da'} \mid \text{al. bol. u rod.} = \text{'ne'},\ \text{al. bol. u dj.} = \text{'da'},\ \text{ITM} \leq 25) = 0,63$ $P(\text{al. dišna bol.} = \text{'da'} \mid \text{al. bol. u rod.} = \text{'ne'},\ \text{al. bol. u dj.} = \text{'ne'},\ \text{ITM} > 25) = 0,07$ $P(\text{al. dišna bol.} = \text{'da'} \mid \text{al. bol. u rod.} = \text{'ne'},\ \text{al. bol. u dj.} = \text{'ne'},\ \text{ITM} \leq 25) = 0,12$

Povezanost pretjerane konzumacije alkohola i alergijskih dišnih bolesti, predloženu modelom, nije bilo moguće ispitati jer ni u jednom izvoru podataka nisu bili mjereni zajedno. Ipak, bilo je moguće ispitati povezanost pretjeranog unosa alkohola i alergijske astme u četvrtom izvoru. Povezanost nije bila značajna: dobiven je omjer šansi 0,33 s 95 %-tnim intervalom pouzdanosti [0,04 – 2,64]. Iako su simptomi depre-

sije i alergijske dišne bolesti bile mjerene zajedno u prvom izvoru, njihovu povezanost nije bilo moguće ispitati iz tog izvora zbog vrlo niske prevalencije simptoma depresije. Kao i kod unosa alkohola, bilo je moguće ispitati samo povezanost simptoma depresije i alergijske astme u četvrtom izvoru. Osobe sa simptomima depresije imale su veće šanse za alergijsku astmu (omjer šansi 2,52), no niti ova povezanost nije bila statistički značajna (95 %-tni interval pouzdanosti [0,93 – 6,85]). Ipak, nedostatatak obje ove analize je što ispitanici s alergijskim rinitisom nisu mogli biti isključeni iz kontrolne skupine zbog nepoznatih podataka o bolesti.

5.2.3.5 Povezanost kontakta sa životinjama s alergijskim dišnim bolestima

Kao što je opisano u poglavlju 4.2.2, povezanost kontakta sa životinjama i alergijskih dišnih bolesti ispitana je samo u ispitanika u gradu. Broj ispitanika iz grada u kontaktu sa životinjama, prema kategorijama ljubimaca (pas, mačka, ptica i glodavac), prikazan je u Tablici 21. Kao i u cijelom uzorku, ispitanici u gradu najviše su u kontaktu bili s mačkama i psima, dok su ptice i glodavci bili manje zastupljeni. Za svaku kategoriju kućnog ljubimca, Tablica 21 prikazuje i usporedbu prevalencija alergijskih dišnih bolesti između ispitanika u redovitom kontaktu s tim ljubimcem i ostalih ispitanika.

Tablica 21: Kućni ljubimci u ispitanika u gradu za koje je bilo poznato posjeduju li ljubimca te imaju li al. dišnu bolest. N: broj ispitanika, p: p-vrijednost testa nezavisnosti (χ^2-test ili Fisherov test).

		1. izvor			2. izvor			3. izvor		
Ljubimac		N	Al. dišna bolest (%)	p	N	Al. dišna bolest (%)	p	N	Al. dišna bolest (%)	p
Pas	da	141	34 (24 %)	$0,758$	30	7 (23 %)	$0,455$	54	3 (6 %)	$0,575$
	ne	181	41 (23 %)		119	36 (30 %)		174	15 (9 %)	
Mačka	da	99	30 (30 %)	$p: 0,039^*$	31	10 (32 %)	$0,634$	41	0 (0 %)	$p: 0,049^*$
	ne	222	44 (20 %)		118	33 (28 %)		187	18 (10 %)	
Ptica	da	18	4 (22 %)	1	6	0 (0 %)	$0,183$	29	0 (0 %)	$0,139$
	ne	305	72 (24 %)		144	43 (30 %)		199	18 (9 %)	
Glodavac	da	24	9 (38 %)	$0,093$	7	2 (29 %)	1	5	0 (0 %)	1
	ne	299	67 (22 %)		142	41 (29 %)		223	18 (8 %)	

* $p < 0,05$

Statistička povezanost redovitog kontakta s mačkom dobivena je u prvom i trećem izvoru podataka. Dok su u prvom izvoru ispitanici u kontaktu s mačkom imali višu prevalenciju alergijskih dišnih bolesti u odnosu na ostale ispitanike (30 % naspram 20 %), u trećem je izvoru situacija bila obrnuta: nijedan ispitanik s mačkom kao kućnim ljubimcem nije imao alergijsku dišnu bolest, u usporedbi s 10 % ispitanika koji nisu držali mačku kao ljubimca.

Ni u jednom izvoru nije pronađena statistički značajna povezanost između redovitog kontakta sa psom, pticom ili glodavcem i alergijskih dišnih bolesti.

5.2.3.6 Povezanost pasivnog pušenja s alergijskim dišnim bolestima

Povezanost pasivnog pušenja i alergijskih dišnih bolesti ispitana je samo u nepušača, kako je opisano u poglavlju 4.2.2. Broj nepušača koji su pasivno pušili za prvi i drugi izvor prikazan je u Tablici 22. Gotovo svi nepušači iz prvog izvora podataka (njih 96 %) pasivno su pušili, u usporedbi s gotovo polovicom nepušača (50 %) u drugom izvoru podataka.

Tablica 22: Pasivno pušenje u nepušača za koje je bilo poznato da li pasivno puše te imaju li al. dišnu bolest. N: broj ispitanika, p: p-vrijednost testa nezavisnosti (χ^2-test ili Fisherov test).

		1. izvor			2. izvor		
		N	Al. dišna bolest (%)	p	N	Al. dišna bolest (%)	p
Pasivno pušenje	da	273	63 (23 %)	0,466	53	15 (28 %)	0,578
	ne	11	1 (9 %)		54	19 (35 %)	

Iako su u prvom izvoru alergijske dišne bolesti bile nešto češće kod pasivnih pušača nego kod ostalih (23 % naspram 9 %), a u drugom nešto rjeđe (28 % kod pasivnih pušača naspram 35 % kod ostalih), nije utvrđena statistički značajna povezanost pasivnog pušenja i alergijskih dišnih bolesti ni u jednom izvoru.

6 Rasprava

U prvom dijelu ovog poglavlja raspravljeni su rezultati vezani za prvi cilj istraživanja (dogradnja metode stapanja podataka iz više izvora), dok je drugi dio Rasprave ostavljen za rezultate vezane za drugi cilj istraživanja (istraživanje čimbenika rizika za nastanak alergijskih dišnih bolesti na podacima iz četiri epidemiološka izvora

odrasle populacije Hrvatske metodom stapanja podataka).

6.1 Dogradnja metode stapanja podataka iz više izvora

U ovom je istraživanju predstavljena i evaluirana nova metoda za procjenu parametara kad podaci iz različitih izvora dijele samo dio čimbenika (podaci s djelomično preklapajućim čimbenicima). Prema saznanjima autorice, ovo je prvi rad u kojem je metoda za procjenu parametara iz podataka s djelomično preklapajućim čimbenicima primijenjena u dogradnji metode za procjenu strukture Bayesove mreže te ujedno i prva epidemiološka primjena dograđene metode.

Metoda stapanja podataka iz više izvora zasad je razvijena za slučaj kad pretpostavljamo da podaci dolaze iz iste populacije, s istim uzročno-posljedičnim vezama među varijablama i njihovim jačinama (koje se, doduše, u dobivenim uzorcima mogu razlikovati zbog pogreške uzorkovanja), te kad su sve varijable kategorijske. Ako se smatra da podaci dolaze iz različitih populacija, s drugačijim uzročno-posljedičnim mehanizmima, tada ni nema zajedničkog kauzalnog modela koji bi se mogao otkriti te bi trebalo analizirati svaki izvor podataka zasebno. Kontinuirane varijable mogu se upotrijebiti prilikom primjene metode za učenje strukture, kao i prilikom primjene metode za učenje parametara. Međutim, prema saznanjima autorice, trenutno nije moguće na jednostavan način, bez dodatnog programiranja u specijaliziranom softveru, izračunati vjerojatnost prisutnosti bolesti u ovom slučaju.

Proces stapanja podataka iz više izvora nije automatiziran te je još uvijek u velikoj mjeri potrebna pomoć stručnjaka ili vanjskog znanja. Prije početka modeliranja potrebo je ocijeniti može li se primijeniti hipoteza da su podaci iz iste populacije, tj. očekuju li se iste uzročno-posljedične veze u svim podacima te očekuju li se iste jačine ovih veza (unutar pogreške nastale zbog upotrebe razliitih uzoraka). Proces

učenja uzročno-posljedičnih veza uvelike olakšava znanje o vezama ili smjerovima koji su sigurno prisutni te o vezama ili smjerovima koji nisu dozvoljeni (primjerice, logički nemogući smjerovi). Nadalje, potrebno je procijeniti u kojoj su mjeri podaci pristrani te koji se njihovi dijelovi mogu koristiti u procjeni. Naposljetku, nakon dobivanja zajedničkog modela, potrebno je odrediti koje se veze zadržavaju za procjenu te što je uzrok, a što posljedica kod povezanosti kod kojih model dopušta više mogućnosti smjera povezanosti.

U nastavku je detaljnije raspravljen svaki od koraka u procesu stapanja podataka (procjena parametara kauzalnog modela, procjena strukture kauzalnog modela te epidemiološka primjena).

6.1.1 Procjena parametara

Metoda za procjenu parametara G-MA, u ovoj disertaciji korištena u dogradnji modela Bayesove mreže, predstavljena je i evaluirana na stvarnim podacima iz epidemioloških meta-analiza u radu autorica Kovačić i Varnai (67), dok je u ovom radu evaluirana na računalno generiranim podacima.

U procjeni parametara, rezultati jasno upućuju na prednost korištenja svih dostupnih izvora podataka u odnosu na korištenje samo onih izvora koji posjeduju potpune podatke (uključujući korištenje samo jednog izvora podataka), kad god je procjena parametara G-MA metodom tehnički moguća. U odnosu na korištenje samo izvora s potpunim podacima (CC-MA metoda), G-MA metoda pokazala je manju pristranost i uže intervale pouzdanosti, što omogućava točniju procjenu rizika i raspona u kojem se nalazi. U stvarnom epidemiološkom primjeru iz rada (67), omjer šansi dobiven G-MA metodom manje se razlikovao od omjera šansi dobivenog iz potpunih podataka nego omjer šansi dobiven CC-MA metodom (maksimalna razlika od 0,06 za G-MA u usporedbi s maksimalnom razlikom od 0,31 za CC-MA). U eksperimentima s računalno generiranim podacima (poglavlje 5.1.1), odstu-

panje CC-MA metode od stvarne vrijednosti procjenjivanog parametra, mjereno srednjom kvadratnom pogreškom, bilo je veće od odstupanja G-MA metode: srednja kvadratna pogreška CC-MA metode tako je u svim eksperimentima bila dva do četiri puta veća od srednje kvadratne pogreške G-MA metode. Nadalje, u primjerima epidemioloških meta-analiza iz rada (67), CC-MA metoda imala je 10 % do 64 % šire intervale pouzdanosti nego G-MA metoda. Slično, u eksperimentima s računalno generiranim podacima, intervali pouzdanosti CC-MA metode bili su u prosjeku 29 % do 77 % širi od intervala pouzdanosti G-MA metode.

G-MA metoda ima ograničenja, koja su detaljnije diskutirana u radu autorica Kovačić i Varnai (67). Mogući su tehnički problemi u procjeni (primjerice, neuspješna konvergencija ili predugo vrijeme procjene), kakvi su uočeni u jednom epidemiološkom primjeru u navedenom radu te u 7 % slučajeva u eksperimentima s računalno generiranim podacima (poglavlje 5.1.1). U takvim je slučajevima moguće rješenje pojednostavljenje grafičkog modela korištenog u procjeni. Pojednostavljenjem se gubi dio podataka, što je sub-optimalno u odnosu na korištenje svih dostupnih podataka, ali opet je bolje od korištenja samo onih izvora s potpunim podacima. Nadalje, metoda je jednostavno primjenjiva samo kad čimbenici slijede određeni uzorak (takozvani monotoni uzorak, Tablica 23), kakav je bio prisutan u većini epidemioloških primjera iz rada (67) te u svim primjerima u ovoj disertaciji. U ostalim slučajevima potrebno je provesti dodatne eksperimente kako bi se provjerila točnost metode ili naučiti strukturu grafičkog modela koji se koristi u procjeni iz podataka, što zahtijeva dodatan trud. Moguće je rješenje također odbaciti dio podataka kako bi se dobio monotoni uzorak. Drugo je moguće rješenje korištenje prethodno predloženih metoda za procjenu parametara iz podataka s djelomično preklapajućim čimbenicima, kao što su metode višestrukog umetanja autora Resche-Rigona i suradnika (36) ili Jolanija i suradnika (37).

G-MA metoda ne rješava problem sporadično nepotpunih podataka (primjerice,

Tablica 23: (a) Monotoni uzorak nepotpunih podataka: bilo koji redak sadrži sve varijable mjerene u retcima ispod. (b) Nemonotoni uzorak: nikakvim preslagivanjem redaka i stupaca ne može se dobiti monotoni uzorak.

(a)

Uzorak	V_1	V_2	V_3
1	✓	✓	✓
2	✓	✓	
3	✓	✓	
4	✓		

(b)

Uzorak	V_1	V_2	V_3
1	✓	✓	✓
2	✓	✓	
3	✓		
4	✓		✓

kad unutar istog istraživanja, tj. izvora podataka, dio ispitanika ne odgovori na određeno pitanje). Takvi bi se problemi trebali rješavati drugim tehnikama prije primjene G-MA metode (primjerice, višestrukim umetanjem za svaki dostupni izvor podataka posebno (35)).

Prednost je G-MA metode to što, za razliku od metoda višestrukog umetanja podataka s djelomično preklapajućim čimbenicima, može uključiti i sumarne podatke iz objavljenih istraživanja (primjerice, omjer šansi i njegovu standardnu pogrešku iz ranije objavljenog znanstvenog rada) i koristiti ih u procjeni ravnopravno s dostupnim mikro-podacima (tablice koja sadrži podatke o vrijednostima svakog mjerenog čimbenika za svakog ispitanika iz istraživanja). Na taj bi se način u epidemiološkoj primjeni mogli uključiti i rezultati drugih istraživanja s istom tematikom u hrvatskoj populaciji. Prije ovakvog združivanja podataka potrebno je ocijeniti, s obzirom na karakteristike ispitanika i vremenske periode obuhvaćene različitim istraživanjima, očekuju li se isti učinci u svakoj sub-populaciji te prema tome odabrati model s fiksnim efektima ili model sa slučajnim efektima.

6.1.2 Procjena strukture

Za razliku od procjene parametara, prednosti učenja strukture kauzalnog modela (uzročno-posljedičnih veza među čimbenicima) iz više izvora naspram učenju iz

samo jednog izvora podataka manje su jasne. Naime, kako su pokazali rezultati iz poglavlja 5.1.2, analiza jednog izvora podataka i analiza četiri izvora podataka pokazale su slične prediktivne vrijednosti i specifičnost, i to i u prepoznavanju povezanosti i njihovih orijentacija (razlikovanja uzroka od posljedice). Naime, za obje vrste analize medijani prediktivnih vrijednosti kod prepoznavanja povezanosti kretali su se u rasponu od 0,80 do 0,86, a medijani prediktivnih vrijednosti kod prepoznavanja orijentacija u rasponu od 0,45 do 0,47. S druge strane, analiza jednog izvora podataka pokazala je višu osjetljivost. U prepoznavanju povezanosti, medijani osjetljivosti u analizi jednog izvora podataka bili su viši za 0,09 i 0,11 od medijana osjetljivosti u analizi četiri izvora podataka. Slično, u prepoznavanju orijentacija povezanosti, medijani osjetljivosti analize jednog izvora podataka bili su viši za 0,13 i 0,17 od medijana osjetljivosti analize četiri izvora podataka. Niža osjetljivost analize četiri izvora podataka posljedica je toga što ova vrsta analize dio povezanosti označava samo kao moguće prisutne te takve povezanosti ne ulaze u izračun osjetljivosti kao točno prepoznate.

Prednost analize sva četiri izvora podataka je što sintetizira znanje iz svih izvora u jednom modelu (grafu). Također, može indicirati moguće povezanosti među varijablama koje ni u jednom izvoru nisu mjerene zajedno te tako postaviti hipotezu za novo istraživanje. Poznato je da model Bayesove mreže ne može zaista razlikovati uzrok od posljedice. Naime, kako je već istaknuto, u slučaju dvije povezane varijable A i B jednaku potporu u podacima imaju hipoteze da A uzrokuje B, da B uzrokuje A ili pak da su A i B povezane putem treće varijable. To se vidi i u prethodno navedenim pozitivnim prediktivnim vrijednostima kod prepoznavanja orijentacija: u barem 50 % ponavljanja eksperimenata više od polovice prepoznatih povezanosti bilo je netočno orijentirano (medijani pozitivnih prediktivnih vrijednosti 0,47 i manje). Ova opaska naglašava važnost intervencije stručnjaka iz područja prilikom izgradnje i interpretacije ovakvih modela, i to i u izgradnji i interpretaciji modela

Bayesove mreže za svaki izvor zasebno, i u odabiru i interpretaciji završnog modela.

S druge strane, modeli su se pokazali relativno dobrima u prepoznavanju povezanosti, i kod analize jednog i kod analize sva četiri izvora podataka. Većina prepoznatih veza bila je točna: kad su izvori podataka bili iz istih populacija, u barem 50 % ponavljanja eksperimenata barem je 84 % prepoznatih povezanost zaista bilo prisutno (medijani pozitivne prediktivne vrijednosti od 0,84 i 0,86), dok je u gotovo svim ponavljanjima barem 60 % prepoznatih povezanosti bilo točno (2,5 %-tni kvantili od 0,60). Kad su izvori podataka bili iz različitih populacija, ove su vrijednosti bile nešto niže, ali još uvijek visoke (medijani pozitivnih prediktivnih vrijednosti od 0,80 i 0,81 te 2,5 %-tni kvantili od 0,60 i 0,56). Međutim, velik udio stvarno prisutnih povezanosti ostajao je neprepoznat: tako je u barem 50 % ponavljanja najmanje oko 40 % povezanosti ostalo neprepoznato (medijani osjetljivosti od 0,47 do 0,61), a u gotovo svim ponavljanjima barem oko 10 % povezanosti nije bilo prepoznato (97,5 %-tni kvantili osjetljivosti u rasponu od 0,72 do 0,92). I iz ovog aspekta je stoga vrlo bitno uključiti stručnjaka iz područja u proces izgradnje modela kako bi uočio povezanosti koje nisu prepoznate modelom.

U epidemiološkoj primjeni izgradnje kauzalnog modela za alergijske dišne bolesti očekivan je i veći broj neprepoznatih povezanosti, s obzirom na veličine uzorka koje su za tri od četiri izvora podataka bile manje nego u eksperimentima. S druge strane, očekivan je i manji udio netočno orijentiranih povezanosti s obzirom da u procesu izgradnje modela i odabira završnog modela logički nemoguće orijentacije nisu dozvoljene.

6.1.3 Evaluacija na testnom skupu

U evaluaciju na epidemiološkim izvorima podataka uključene su tri vrste analize: 1) metoda analize djelomično preklapajućih podataka iz više izvora, odnosno Bayesova mreža četiri izvora, 2) Bayesova mreža jednog izvora te 3) logistička regresija

jednog izvora. U istraživanju čimbenika rizika za nastanak alergijskih dišnih bolesti niti jedna vrsta analize nije bila najbolja s obzirom na sve promatrane parametre (osjetljivost, specifičnost, točnost i prediktivne vrijednosti).

Model Bayesove mreže četiri izvora pokazao je najbolji omjer osjetljivosti i specifičnosti. Naime, imao je najveću površinu ispod ROC krivulje u dva od tri slučaja (0,63 na testnom skupu prvog izvora te 0,74 na testnom skupu drugog izvora, u usporedbi s 0,62 i 0,46 za model Bayesove mreže jednog izvora i 0,51 i 0,56 za model logističke regresije za iste izvore podataka). S druge strane, model logističke regresije pokazao je najbolju točnost i najbolji omjer pozitivne i negativne prediktivne vrijednosti. Najbolju postignutu točnost imao je u dva od tri slučaja (0,85 na testnom skupu drugog izvora i 0,98 na testnom skupu drugog izvora, u usporedbi s 0,79 i 0,95 za model Bayesove mreže jednog izvora i 0,82 i 0,95 za model Bayesove mreže četiri izvora). Najbolji omjer pozitivne i negativne vrijednosti pokazao je također u dva od tri slučaja, za testni skup prvog i trećeg izvora. U prvom izvoru pozitivna i negativna prediktivna vrijednost logističke regresije bile su 1 i 0,78. Iako je Bayesova mreža jednog izvora imala malo veću negativnu prediktivnu vrijednost za ovaj izvor podataka (0,80), njena pozitivna prediktivna vrijednost bila je bitno lošija od vrijednosti logističke regresije (0,75). Za Bayesovu mrežu četiri izvora bilo je moguće izračunati samo negativnu prediktivnu vrijednost u ovom slučaju i ona je iznosila 0,77. U testnom skupu trećeg izvora pozitivna i negativna prediktivna vrijednost logističke regresije bile su redom 1 i 0,98, u usporedbi s negativnim prediktivnim vrijednostima Bayesovih mreža koje su iznosile 0,95 (pozitivne prediktivne vrijednosti nije bilo moguće izračunati u ovom slučaju).

U usporedbi samo s Bayesovom mrežom jednog izvora, Bayesova mreža četiri izvora imala je malo bolje rezultate. Naime, imala je dosta bolji omjer osjetljivosti i specifičnosti u dva od tri slučaja (površina ispod krivulje bila je 0,74 i 0,82 za drugi i treći izvor podataka, u usporedbi s 0,46 i 0,54 za mrežu jednog izvora), a

bolji omjer imala je i u trećem slučaju (prvi izvor podataka), iako je razlika bila mala (0,63 za mrežu četiri izvora i 0,62 za mrežu jednog izvora). S druge strane, po pitanju točnosti i prediktivnih vrijednosti ova dva modela bila su podjednaka. U prvom izvoru model jednog izvora imao je više sve navedene vrijednosti, dok je model četiri izvora najveću točnost postigao za model koji sve ispitanike klasificira kao zdrave. U drugom izvoru podataka bilo je obrnuto, a u trećem su oba modela najveću točnost postigla za model koji sve ispitanike klasificira kao zdrave te su imala jednake vrijednosti točnosti i prediktivnih vrijednosti.

Zanimljivo je da je model četiri izvora najveću prednost u odnosu na modele jednog izvora podataka pokazao u drugom izvoru podataka, gdje je uzorak bio najmanji (165 ispitanika). Naime, model četiri izvora imao je bitno veću površinu ispod ROC krivulje u ovom izvoru od ostala dva modela (0,74 u usporedbi s 0,56 i 0,46 za modele jednog izvora). Iako je imao malo nižu točnost od modela logističke regresije, razlika nije bila velika (0,82 u usporedbi s 0,85), a omjer prediktivnih vrijednosti bio je bolji nego isti omjer kod modela logističke regresije (pozitivna i negativna prediktivna vrijednost 1 i 0,81, u usporedbi s bitno nižom pozitivnom prediktivnom vrijednosti od 0,75 te malo višom negativnom prediktivnom vrijednosti od 0,86 za model logističke regresije).

Više je mogućih razloga zašto se model četiri izvora nije pokazao najboljim prema svim promatranim parametrima u ostala dva izvora. Naime, pretpostavka metode za procjenu strukture je da su svi izvori podataka dobiveni iz populacija u kojima vrijede iste uzročno-posljedične veze među odabranim čimbenicima. S obzirom na razlike u vremenskim periodima u kojima su istraživanja provedena te razlike u dobi i radnim mjestima ispitanika, moguće je da navedena pretpostavka nije bila zadovoljena. Iz istog je razloga moguće i da se razlika u jačinama uzročno-posljedičnih veza u sub-populacijama iz različitih istraživanja ne može objasniti samo pogreškom uzorkovanja te da se radi zapravo o različitim populacijama. Takve bi podatke tre-

balo analizirati prikladnijim modelom sa slučajnim efektima, što ovdje nije bilo izvedivo zbog premalog broja izvora podataka. Mali broj izvora podataka te mala veličina uzorka drugog izvora podataka, kao i niska prevalencija slučajeva bolesti u trećem izvoru podataka, također su mogli negativno utjecati na točnost procjene parametara. Također, moguće je da je bila prisutna pristranost izbora (engl. *selection bias*). Naposljetku, iz modela četiri izvora dobivenog metodom COmbINE za procjenu je odabran samo jedan mogući model, prilagođen onome što je bilo moguće procijeniti s obzirom na dostupnost čimbenika u različitim izvorima. Moguće je da bi model koji bi uključio iz tog razloga izostavljene povezanosti dao bolje rezultate.

Na temelju samo ovih rezultata, zaključili bismo da ćemo za predviđanje prisutnosti alergijske dišne bolesti koristiti model logističke regresije ako su nam najvažnije točnost i prediktivne vrijednosti, a model Bayesove mreže četiri izvora ako su nam važnije osjetljivost i specifičnost. Model Bayesove mreže četiri izvora preferiran je u odnosu na isti model samo jednog izvora. Također, model više izvora podataka preferiran je u odnosu na oba modela jednog izvora podataka u slučaju malog uzorka.

Ipak, na temelju provedene evaluacije, teško je donijeti pouzdane zaključke. Evaluacija je provedena na samo jednom odabiru skupa za treniranje i skupa za testiranje. Za pouzdanije zaključke trebalo bi provesti barem nekoliko desetaka ovakvih evaluacija. Nažalost, zbog vremenski zahtjevnog postupka izgradnje modela iz više izvora, pri čemu je samo mali dio postupka automatiziran, to u ovom istraživanju nije bilo moguće. Također, bilo bi poželjno provesti ovakvu evaluaciju i na drugim izvorima iz istraživanja alergijskih dišnih bolesti u hrvatskoj populaciji, no, s obzirom na nedostupnost podataka, ni to nije bilo moguće.

Budući da se niti jedna metoda za analizu podataka nije pokazala najboljom prema svim ispitivanim mjerama valjanosti, u drugom dijelu Rasprave (istraživanje čimbenika rizika za nastanak alergijskih dišnih bolesti na podacima iz četiri epidemi-

ološka izvora odrasle populacije Hrvatske) komentirani su rezultati svih analiza.

6.2 Istraživanje čimbenika rizika za nastanak alergijskih dišnih bolesti

U ovom dijelu Rasprave detaljnije su razmotreni dobiveni rezultati povezanosti čimbenika uključenih u ovo istraživanje s prisutnošću alergijskih dišnih bolesti.

Pregled čimbenika prema povezanosti s prisutnošću alergijske dišne bolesti dan je u Tablici 24. Od promatranih 18 čimbenika, devet ih nije bilo značajno izravno povezano s bolešću ni u jednom izvoru podataka (dob, obrazovanje, mjesto stanovanja, broj soba po osobi, centralno grijanje, unos voća i povrća, zadovoljstvo zdravljem, neuroticizam i pasivno pušenje). Od ostalih čimbenika, samo su alergijske bolesti u djetinjstvu pokazale značajnu povezanost s prisutnošću alergijske dišne bolesti u svim promatranim modelima (logistička regresija i Bayesova mreža trećeg izvora podataka te zajednički model sva četiri izvora podataka). Alergijske bolesti u roditelja bile su povezane s prisutnošću bolesti u dva od tri izvora podataka i zajedničkom modelu, indeks tjelesne mase u jednom od tri izvora podataka i zajedničkom modelu, dok su spol, aktivno pušenje i tjelesna aktivnost bile značajno povezane s prisutnošću bolesti samo u jednom izvoru podataka. Povezanost simptoma depresije i konzumacije alkohola s prisutnošću bolesti nije bilo moguće ispitati u prva tri izvora, ali moguća povezanost naznačena je u zajedničkom modelu sva četiri izvora podataka. Od kućnih ljubimaca, značajna povezanost s alergijskim dišnim bolestima dobivena je kod ispitanika u kontaktu s mačkama u dva izvora podataka. U nastavku su detaljnije diskutirane ove povezanosti.

[3]U matematičkom smislu, kako je objašnjeno u poglavlju 4.2.1.5.

Tablica 24: Čimbenici značajno povezani s prisutnošću alergijske dišne bolesti.

Čimbenik	1. izvor	2. izvor	3. izvor	4 izvora
Muški spol	LR ↑			
Dob				
Aktivno pušenje			LR ↓, BN ↓	
Indeks tj. mase	LR ↓, BN ↓			sigurno[3]
Obrazovanje				
Mjesto stanovanja				
Broj soba po osobi				
Centralno grijanje				
Al. bolesti u roditelja	LR ↑, BN ↑		LR ↑	sigurno
Al. bolesti u djetinjstvu			LR ↑, BN ↑	moguće
Dovoljna tj. aktivnost			LR ↑	
Unos voća i povrća				
Zadovoljstvo zdravljem				
Neuroticizam				
Simptomi depresije				moguće
Konzumacija alkohola				moguće
Kućni ljubimci	χ^2/Fisher ↑		χ^2/Fisher ↓	
Pasivno pušenje				

Oznake: povećanje šansi ili vjerojatnosti prisutnosti bolesti označeno je simbolom ↑, a smanjenje simbolom ↓. Siva boja označava varijablu koja nije bila uključena u model za pripadni izvor podataka. LR: višestruka logistička regresija; BN: Bayesova mreža (izravan uzrok ili izravna posljedica alergijske dišne bolesti); χ^2/Fisher: χ^2–test ili Fisherov test.

6.2.1 Povezanost alergijskih dišnih bolesti i alergijskih bolesti u djetinjstvu

Čimbenik koji je bio najsnažnije povezan s prisutnošću alergijskih dišnih bolesti bile su alergijske bolesti u djetinjstvu ispitanika. Omjer šansi povezan s alergijskim bolestima u djetinjstvu bio je daleko viši od omjera šansi za ostale čimbenike: omjer šansi bio je 13,28 s 95 %-tnim intervalom pouzdanosti [4,63 – 38,10] u jednostrukoj logističkoj regresiji, odnosno 28,15 [6,59 – 120,31] u višestrukoj regresiji (izvor podataka: prethodni pregledi u farmaceutskoj industriji). Osim u logističkoj regresiji, i u modelu Bayesove mreže izdvojen je kao čimbenik izravno povezan s alergijskim

dišnim bolestima. U modelu Bayesove mreže sva četiri izvora također je bio izravno povezan s alergijskim dišnim bolestima, iako je ostavljena i mogućnost da postoji čimbenik koji nije uključen u postojeće čimbenike koji bi bio izravni uzrok i alergijskih bolesti u djetinjstvu i alergijskih dišnih bolesti kasnije u životu.

Snažna povezanost alergijskih bolesti u djetinjstvu i alergijskih dišnih bolesti kasnije u životu u skladu je s ranijim istraživanjima. Tako su autori Grad i Morgan u preglednom radu o prognozi astme i zviždanja u prsima u djetinjstvu zaključili da se simptom ranog zviždanja u prsima (engl. *wheezing*) koji traje barem do treće godine života u pravilu zadržava i kasnije u životu, sve do odrasle dobi (82). Slično, Schoenwetter u preglednom radu o tijeku alergijskog rinitisa ističe da alergijski rinitis koji započne u djetinjstvu ili adolescenciji u velikom postotku ostaje prisutan i u odrasloj dobi (15): preko 80 % djece s alergijskim rinitisom i nakon 10 godina još uvijek ima simptome (83). Također, smatra se da se u 40 % onih koji u odrasloj dobi imaju alergijski rinitis simptomi pojave do šeste godine života (84).

U istraživanjima koja su pratila ispitanike od djetinjstva do odrasle dobi zaključeno je da alergijske bolesti u djetinjstvu mogu povećati rizik za pojavu astme kasnije u životu: Burgess i suradnici su, prateći preko 5000 djece od sedme godine života, zaključili da je alergijski rinitis u dobi do sedme godine života značajno povezan s povećanjem rizika za pojavu astme i u dobi od 13. do 20. godine i u dobi od 21. do 44. godine (85). Koristeći uzorak iz istog istraživanja, autori su zaključili da je i ekcem u ranom djetinjstvu (u dobi do sedme godine života) povezan sa značajnim povećanjem rizika za astmu i u adolescenciji i u odrasloj dobi (86). Slične zaključke u istraživanju praćenja dobili su Toelle i suradnici: atopija, alergijski rinitis, ekcem, zviždanje u prsima i dijagnoza astme u dobi od osam do deset godina bili su značajno povezani sa simptomima astme 15 – 17 godina kasnije (87).

Iako su povećane šanse za alergijske dišne bolesti u ispitanika s alergijskim bolestima u djetinjstvu konzistentne s navedenim istraživanjima, treba imati na umu

da nije bilo moguće potpuno pouzdano procijeniti pripadni omjer šansi te da je on u stvarnosti u populaciji vjerojatno nešto manji. Naime, alergijske bolesti u djetinjstvu zabilježene su u samo jednom izvoru (registar prethodnih pregleda u farmaceutskoj industriji) u kojem je bilo malo ispitanika s alergijskim dišnim bolestima (21 ispitanik, 7 %). Poznato je da u takvim situacijama logistička regresija precjenjuje omjere šansi veće od 1 (88). Niska prevalencija alergijskih dišnih bolesti uzrok je i širokih intervala pouzdanosti u logističkoj regresiji za čimbenike u ovom izvoru podataka, uključujući interval pouzdanosti za alergijske bolesti u djetinjstvu. Naposljetku, budući da je ovo istraživanje bilo retrospektivno, moguće je da je došlo i do pristranosti zbog prisjećanja (engl. *recall bias*). Naime, nedavno istraživanje pokazalo je da se atopijskog dermatitisa u djetinjstvu bolje prisjećaju ispitanici koji trenutno imaju atopijski dermatitis, rinitis ili astmu te ispitanici koji su imali dugotrajni dermatitis (preko deset godina) u djetinjstvu (89). Stoga bismo očekivali da su se i u ovom istraživanju ispitanici s trenutnim atopijskim dišnim tegobama bolje prisjećali tegoba u djetinjstvu nego ostali ispitanici, što je dodatno povećalo procijenjeni omjer šansi.

6.2.2 Povezanost alergijskih dišnih bolesti i pozitivne obiteljske anamneze

Dobro je poznato da su alergijske dišne bolesti snažno povezane i s obiteljskim naslijeđem, tj. da se češće pojavljuju u obiteljima s poviješću alergijskih bolesti (2). Kako Halken i Host navode u svom preglednom radu, iz prethodnih je istraživanja poznato da samo oko 10 % djece iz obitelji bez povijesti atopijskih bolesti razvije atopijsku bolest, u usporedbi s 20 % do 30 % djece kojima jedan roditelj (ili brat/sestra) ima atopijsku bolest i 50 % djece kojima oba roditelja imaju atopijsku bolest (90). Burke i suradnici su u preglednom radu o povezanosti atopijskih bolesti u obitelji

i dječje astme zaključili da pozitivna obiteljska anamneza značajno povećava šanse za astmu u djetinjstvu, pri čemu su omjeri šansi istraživanja uključenih u pregled varirali od 1,7 do 6,8 (91). U drugom istraživanju s odraslim ispitanicima, ispitanici s pozitivnom obiteljskom anamnezom astme ili rinitisa imali su tri do četiri puta veći rizik za astmu te dva do šest puta veći rizik za rinitis u odnosu na ispitanike bez atopijskih bolesti u obitelji (92).

Navedeni su rezultati u skladu s rezultatima ove disertacije: u sva tri izvora podataka u kojima je bila poznata obiteljska anamneza, ispitanici s pozitivnom obiteljskom anamnezom imali su barem dvaput veće šanse za alergijsku dišnu bolest u odnosu na ostale ispitanike (omjeri šansi u rasponu od 2,03 do 6,84 u jednostrukim analizama te od 1,97 do 5,83 u modelima višestruke regresije). Pritom su dobiveni omjeri šansi u prvom i trećem izvoru bili statistički značajni, za razliku od omjera šansi u drugom izvoru. Nedostatak statističke značajnosti u drugom izvoru može se objasniti nedostatkom statističke snage zbog malog uzorka. Iako su alergijske bolesti u roditelja bile povezane s alergijskim dišnim bolestima u zajedničkoj analizi sva četiri izvora podataka, povezanost, osim u drugom izvoru podataka, nije nađena niti u Bayesovoj mreži za treći izvor podataka, i to usprkos statistički značajnom visokom omjeru šansi u modelu logističke regresije. Kao što je već spomenuto u diskusiji o metodološkom dijelu ove disertacije, korišteni algoritam za Bayesove mreže prepoznaje tek oko 60 % stvarno prisutnih povezanosti te je iz tog razloga očekivano i da neke važne povezanosti, poput ove, budu ispuštene iz modela. Također, s obzirom na literaturno znanje, iznenađujuće je i da model Bayesove mreže za treći izvor nije prepoznao vezu između alergijskih bolesti u roditelja i alergijskih bolesti u djetinjstvu. Ipak, iako su ispitanici s pozitivnom obiteljskom anamnezom u ovom izvoru zaista imali oko dva i pol puta veće šanse za alergijske bolesti u djetinjstvu (omjer šansi 2,60 s 95 %-tnim intervalom pouzdanosti [0,81 – 8,42]), povezanost nije bila statistički značajna, i to zbog velike standardne pogreške uslijed niskih prevalencija

alergijskih bolesti u roditelja (28 ispitanika, 9%) i alergijskih bolesti u djetinjstvu (20 ispitanika, 7 %) u ovom izvoru.

Ove opaske dodatno naglašavaju potrebu za uključivanjem stručnjaka, odnosno, vanjskog znanja, u proces izrade kauzalnog modela, pogotovo u slučaju malih uzoraka.

6.2.3 Spolne razlike u prevalencijama alergijskih dišnih bolesti

Veće šanse za alergijske dišne bolesti kod muškaraca zabilježene su u logističkim regresijama za sva tri izvora podataka (omjeri šansi u rasponu od 1,18 do 1,39 u jednostrukim regresijama te u rasponu od 1,84 do 2,09 u višestrukim regresijama), iako je razlika među spolovima bila statistički značajna samo u prvom izvoru podataka (studenti preddiplomskih studija). U Bayesovim mrežama nije zabilježen izravan utjecaj spola na alergijske dišne bolesti.

Spolne razlike u prevalenciji astme opisane su u brojnim istraživanjima (93, 94, 95, 96). U dječjoj dobi astma je učestalija u dječaka u usporedbi s djevojčicama. U adolescenciji se ovaj odnos počinje mijenjati i prevalencija se izjednačava između spolova, dok je u odrasloj dobi astma učestalija u žena u odnosu na muškarce. Ovi se podaci odnose na ukupnu prevalenciju astme, koja uključuje i atopijsku i neatopijsku astmu. Čini se, međutim, da je u odraslih osoba samo prevalencija neatopijske astme viša u žena u odnosu na muškarce, dok je prevalencija atopijske astme podjednaka u oba spola (94, 97, 98).

Kao i u slučaju astme, čini se da je utjecaj spola drugačiji kod atopijskog i neatopijskog rinitisa (99). Tako je u prospektivnom istraživanju Isle of Wight Birth Cohort Study (100) u djevojaka u dobi od 18 godina uočena veća prevalencija neatopijskog rinitisa u odnosu na mladiće iste dobi, dok je prevalencija atopijskog

rinitisa bila 2,5 puta viša u mladića. Slično, u europskom istraživanju MeDALL nađeno je da se alergijski rinitis rjeđe javlja u djevojaka nego u mladića i u razdoblju nakon puberteta (101). Podaci o prevalenciji rinitisa u talijanskoj populaciji odrasle dobi (20 do 84 godine života) pokazali su višu prevalenciju nealergijskog rinitisa u žena u odnosu na muškarce u dobnoj skupini od 20–44 godine života, te nižu prevalenciju u žena u dobnim skupinama starijim od 44 godine. Nasuprot tome, prevalencija alergijskog rinitisa bila je u svim dobnim skupinama podjednaka u oba spola, odnosno nešto viša u muškaraca u odnosu na žene kad je analiza kontrolirana za potencijalne čimbenike utjecaja (79). Ovi nalazi, međutim, nisu jednoznačni s obzirom da su objavljena i istraživanja koja su pokazala veću prevalenciju simptoma alergijskog rinitisa u žena nego u muškaraca odrasle dobi (npr. (102)), ili pak nisu pokazala spolne razlike u prevalenciji alergijskog rinitisa u odraslih (103).

U epidemiološkim izvorima korištenim u ovom disertacijskom istraživanju vrlo je mali broj ispitanika imao alergijsku astmu bez istodobno prisutnog alergijskog rinitisa: osam ispitanika u prvom izvoru, dvoje u drugom izvoru, te jedan u trećem izvoru (u četvrtom izvoru nepoznat je broj osoba koje su imale samo alergijsku astmu, s obzirom da prevalencija alergijskog rinitisa nije ispitivana). Stoga se može pretpostaviti da je za veću prevalenciju alergijskih dišnih bolesti kao zbirne varijable (koja uključuje i alergijski rinitis i alergijsku astmu) nađenu u muškog spola odgovorna viša prevalencija alergijskog rinitisa u muških ispitanika. U prilog ove pretpostavke govori i rezultat naknadno provedene logističke regresije za prvi izvor (jedini izvor u kojem je spolna razlika za prevalenciju astme bila statistički značajna u modelu logističke regresije), kontrolirane za alergijski rinitis. U ovom modelu, naime, spolna razlika u prevalenciji alergijske astme nije nađena (omjer šansi 0,99 s 95 %-tnim intervalom pouzdanosti [0,43 – 2,31]).

Također, u skladu s brojnim prethodnim istraživanjima (npr. (104, 105, 106, 107)), analiza sva tri izvora u kojima je podatak o atopiji bio prisutan (povišeni

ukupni serumski IgE ili pozitivan kožni test ubodom na standardne inhalacijske alergene) pokazala je višu prevalenciju atopije u muških u odnosu na ženske ispitanike (omjer šansi kretao se u rasponu od 1,69 do 1,99, sa statistički značajnom razlikom za prvi i treći izvor).

6.2.4 Povezanost alergijskih dišnih bolesti i čimbenika životnog stila

U nastavku su diskutirani rezultati analize čimbenika životnog stila (indeks tjelesne mase, izloženost duhanskom dimu, konzumacija alkohola i tjelesna aktivnost). Za razliku od analiza alergijskih bolesti u roditelja i u djetinjstvu, gdje je jasan vremenski slijed promatranih događaja, nedostatak ovih analiza je što ne možemo reći ništa o odnosu između čimbenika i incidencije bolesti u vremenu. Povezanosti bi bilo bolje ispitivati u prospektivnim istraživanjima, no to je nedostatak s kojim se susreću i druga presječna istraživanja.

Odgovore na ova pitanja davali su ispitanici na temelju samoprocjene, što znači da je i ovdje bila moguća pristranost zbog prisjećanja. Također, moguće je da su ispitanici, svjesno ili nesvjesno, željeli ostaviti dojam zdravijeg životnog stila te da su umanjivali svoju konzumaciju alkohola ili težinu, a precjenjivali bavljenje tjelesnom aktivnošću. Ipak, to su sve dobro poznati nedostaci i drugih epidemioloških istraživanja.

6.2.4.1 Indeks tjelesne mase

Indeks tjelesne mase bio je povezan s alergijskim dišnim bolestima u prvom izvoru (studenti preddiplomskih studija), a izravna povezanost dobivena je i u zajedničkom modelu sva četiri izvora podataka. Pritom su u svim analizama, neovisno o statističkoj značajnosti, procijenjeni učinci indeksa tjelesne mase bili slični: ispitanici s

višim indeksom tjelesne mase imali su manje šanse za alergijske dišne bolesti (omjeri šansi u rasponu od 0,86 do 0,91 u modelima višestruke logističke regresije).

Budući da podaci iz literature pokazuju da bi odnos između tjelesne mase i bolesti mogao biti drugačiji za alergijsku astmu i alergijski rinitis, u nastavku su sažeti i diskutirani i rezultati za svaku bolest posebno.

Povezanost indeksa tjelesne mase i astme bilo je moguće ispitati u tri izvora podataka uključena u ovo istraživanje: prvom, drugom i četvrtom izvoru (u trećem izvoru, koji sadrži podatke iz registra prethodnih pregleda u farmaceutskoj industriji, samo je jedan ispitanik imao astmu). U četvrtom izvoru podataka (WHS) osobe s prekomjernim tjelesnom težinom (indeks tjelesne mase od 25 i više) imale su značajno više šanse za alergijsku astmu u odnosu na osobe s normalnom težinom (omjer šansi s pripadnim 95 %-tnim intervalom pouzdanosti bio je 2,65 [1,45 – 4,83]), dok za pothranjene ispitanike analiza nije provedena zbog premalog broja takvih ispitanika. U prva dva izvora podataka (studenti i administrativni radnici) odnos je bio obrnut: osobe s prekomjernom tjelesnom težinom imale su manje šanse za alergijsku astmu (omjer šansi bio je 0,11 [0,00 – 0,83] u prvom izvoru te 0,84 [0,32 – 2,20] u drugom izvoru). Pothranjene su osobe u prvom izvoru imale veće šanse za astmu, iako ne značajno veće (omjer šansi 1,33 [0,46 – 3,27]), dok u druga dva izvora analiza nije uključila pothranjene zbog premalog broja takvih ispitanika.

Alergijski rinitis i indeks tjelesne mase nisu bili povezani: iako su ispitanici s prekomjernom tjelesnom težinom imali manje šanse za rinitis u odnosu na osobe s normalnom težinom (omjeri šansi u rasponu od 0,44 do 0,54), statistička značajnost nije dobivena. Značajna razlika u prevalenciji rinitisa nije dobivena ni u usporedbi pothranjenih ispitanika s onima normalne težine, iako su pothranjeni ispitanici imali malo veće šanse za bolest (omjer šansi 1,19 [0,59 – 2,40] u prvom izvoru podataka).

Dosadašnja istraživanja podržala su prvenstveno hipotezu o povezanosti indeksa tjelesne mase i astme, pri čemu je viši indeks tjelesne mase bio povezan s većim

rizikom od astme. Ovaj je rezultat u skladu s rezultatima analize četvrtog izvora, koji je sadržavao gotovo 1000 ispitanika iz opće populacije.

Povezanost astme i debljine istaknuta je kao zaključak više preglednih radova koji su se bavili ovom tematikom (108, 109, 110, 111). Pritom je u svim navedenim preglednim radovima primijećen trend da stupanj preuhranjenosti utječe na povećanje rizika. Naime, u usporedbi s ispitanicima normalne tjelesne mase, pretili ispitanici (indeks tjelesne mase od 30 i više) imali su veći rizik za astmu nego ispitanici prekomjerne tjelesne težine (indeks tjelesne mase od 25 do 30).

Mehanizam ove povezanosti još nije razjašnjen, iako je predloženo nekoliko hipoteza. S jedne strane, debljina mijenja mehaniku pluća, što može dovesti do suženja dišnih puteva i simptoma karakterističnih za astmu, kao što su otežano disanje i kratak dah. Drugo predloženo objašnjenje je da upalni procesi povezani s debljinom mogu pridonijeti razvoju astme. Također, druge bolesti koje su povezane s debljinom, poput dijabetesa tipa 2 ili povišenog krvnog tlaka, mogu pogoršati simptome astme. Ove su hipoteze detaljno diskutirane u radu autorice Shore (112).

S druge strane, kako se diskutira u radu autora Forda (111), astma bi mogla utjecati na povećanje tjelesne mase utoliko što bi osobe s astmom mogle izbjegavati tjelesni napor i pribjegavati sjedilačkom načinu života. Istaknuto je i da bi uzimanje kortikosteroida u sklopu terapije astme također moglo dovesti do povećanja tjelesne mase kod ovakvih ispitanika. Ipak, ovo objašnjenje čini se vjerojatnim samo za teže slučajeve astme. Također, s obzirom na poznatu vezu astme i depresije, diskutiranu u drugom dijelu ovog poglavlja, predloženo je i da bi povećanje tjelesne težine u astmatičara moglo biti posljedica uzimanja antidepresiva.

Iako smjer povezanosti astme i debljine nije sa sigurnošću utvrđen, zaključak preglednog rada koji je analizirao osam prospektivnih istraživanja u odrasloj populaciji (111) išao je u prilog hipotezi da prekomjerna tjelesna težina prethodi astmi: u sedam od osam uključenih istraživanja dobivena je značajna povezanost između

povišenog indeksa tjelesne mase na početku istraživanja i naknadne pojave astme tijekom perioda praćenja (procijenjeni relativni rizici u rasponu od 1,60 do 3,00). Pritom je u šest od ovih osam istraživanja dobivena pozitivna povezanost stupnja preuhranjenosti i povećanja rizika u osoba povišenog indeksa tjelesne mase.

Nije isključeno ni da su astma i debljina povezane putem trećeg čimbenika koji uzrokuje oba stanja, kao što su zajednički genetički čimbenici (112).

Iako se sama povezanost debljine i astme ne dovodi u pitanje, povezanost debljine i atopije još uvijek je nejasna (113, 114, 115). U nedavnom preglednom istraživanju zaključeno je da je povezanost astme i debljine slabija kod atopijskog tipa astme (108). Nadalje, u nekoliko istraživanja s velikim brojem odraslih ispitanika nije dobivena veza atopije i indeksa tjelesne mase (116, 117, 118), iako postoje i takva istraživanja gdje je povezanost bila značajna (119).

Rezultati o nedostatku povezanosti alergijskog rinitisa i indeksa tjelesne mase dobiveni u ovom istraživanju nisu u kontradikciji s prethodnim istraživanjima. Sličan je rezultat dobiven u velikom istraživanju European Community Respiratory Health Study s oko 15 500 ispitanika, gdje su ispitanici s prekomjernom tjelesnom težinom imali nešto manje šase za alergijski rinitis, iako ne značajno manje: dobiveni omjer šansi za alergijski rinitis bio je 0,87 [0,75 – 1,02] za ispitanike s prekomjernom tjelesnom masom (indeks tjelesne mase u rasponu od 25 do 30) te 0,94 [0,72 – 1,22] za pretile ispitanike (indeks tjelesne mase od 30 i više), u odnosu na ispitanike normalne tjelesne težine (118).

Ipak, dobiveni omjeri šansi za alergijski rinitis za prva tri izvora u ovom istraživanju bili su primjetno niži, a rezultati o značajnoj povezanosti alergijske astme i povišenog indeksa tjelesne mase u prva dva izvora podataka suprotna su zaključcima prethodnih istraživanja. Jedno je moguće objašnjenje da se radi o obrnutoj kauzalnosti. Naime, budući da su sva ova istraživanja bila presječna te da ne možemo znati da li je manji indeks tjelesne mase uzrok ili posljedica alergijske dišne bole-

sti, moguće je da su ispitanici počeli zdravije živjeti nakon pojave bolesti, kako bi umanjili njezine učinke.

U prva tri izvora podataka, gdje su ispitanici s prekomjernom tjelesnom težinom imali manje šanse za alergijske dišne bolesti (omjeri šansi manji od 1 u modelu višestruke regresije), manje šanse za bolest imali su i pušači (omjeri šansi u rasponu od 0,16 do 0,90), ispitanici koji su uzimali nedovoljno voća i povrća (omjer šansi 0,79 u prvom izvoru) te ispitanici koji su se nedovoljno bavili tjelesnom aktivnošću (omjeri šansi za dovoljnu tjelesnu aktivnost 1,07 u prvom izvoru i 3,81 u trećem izvoru podataka). Iako su samo pušenje i tjelesna aktivnost u trećem izvoru pokazali statistički značajnu povezanost, vidljiv je trend da su ispitanici s alergijskim dišnim bolestima imali zdravije stilove života, što bi podržalo hipotezu o obrnutoj kauzalnosti kao jedno moguće objašnjenje ovog trenda.

6.2.4.2 Konzumacija alkohola

Konzumacija alkohola predložena je kao rizični faktor za razvoj alergijskih bolesti, no njihova povezanost još uvijek nije razjašnjena. Pregledi prethodnih istraživanja slažu se da postoji dovoljno dokaza da unos alkohola povećava razinu ukupnih serumskih IgE protutijela (120, 121, 122, 123). Međutim, istraživanja koja su proučavala povezanost konzumacije alkohola s jedne strane te specifičnog IgE-a i rezultata kožnog testa ubodom s druge strane dobila su kontradiktorne rezultate o njihovoj povezanosti (105, 124, 125, 126). U jednom je pak istraživanju konzumacija alkohola bila značajno povezana s alergijskim rinitisom koji traje cijelu godinu (engl. *perennial allergic rhinitis*), ali ne i sa sezonskim rinitisom (127).

Povezanost alergijskih dišnih bolesti i konzumacije alkohola predložena je kao moguća u zajedničkom modelu sva četiri izvora. Budući da ove dvije varijable nisu bile mjerene zajedno ni u jednom izvoru, ovu povezanost nije bilo moguće ispitati te je ispitana samo povezanost alergijske astme i prekomjerne konzumacije alkohola u

četvrtom izvoru podataka (WHS). Ispitanici s prekomjernim unosom alkohola imali su manje šanse za alergijsku astmu, iako povezanost nije bila značajna (omjer šansi 0,33 s 95 %- tnim intervalom pouzdanosti [0,04 – 2,64]).

Slično kao i kod indeksa tjelesne mase, niža konzumacija alkohola kod ispitanika s astmom mogla bi se objasniti njihovom željom da spriječe pogoršanje simptoma. Naime, poznato je da konzumacija alkohola može dovesti do bronhokonstrikcije u astmatičara: prema prethodnim istraživanjima, alergijsku reakciju ili reakciju sličnu alergijskoj nakon konzumacije alkohola iskusi barem 30 % astmatičara u bjelačkoj populaciji (121, 122). Isti je rezultat dobiven u istraživanju autorice Assing i suradnika, gdje su ispitanici sa simptomima astme i ispitanici sa simptomima rinitisa značajno manje konzumirali alkohol od ispitanika bez alergijskih dišnih bolesti (126).

6.2.4.3 Izloženost duhanskom dimu

Kao i za alkohol, postoje istaživanja koja su pokazala da je pušenje povezano s većim vrijednostima ukupnih serumskih IgE protutijela (120). Iako, kako se navodi u nedavnom preglednom radu autora Yooa i Perzanowskog, neka istraživanja podupiru tvrdnju da bi izloženost duhanskom dimu mogla posredno utjecati na alergijsku senzibilizaciju (primjerice, tako što čini epitel propusnijim za alergene), rezultati o povezanosti pušenja i senzibilizacije na alergene iz epidemioloških istraživanja bili su kontradiktorni (128).

U ovom istraživanju povezanost pušenja i alergijskih dišnih bolesti bilo je moguće ispitati u prva tri izvora podataka (studenti, administrativni radnici i registar prethodnih pregleda u farmaceutskoj industriji). U sva tri izvora povezanost je bila negativna: pušači su imali manje šanse za alergijsku dišnu bolest (omjeri šansi u rasponu od 0,31 do 0,95 u jednostrukim logističkim regresijama, odnosno u rasponu od 0,16 do 0,90 u modelima višestruke regresije). Pritom je samo u trećem izvoru podataka pušenje bilo značajno povezano s bolesti, i u regresijskom modelu i u mo-

delu Bayesove mreže. Pasivno je pak pušenje ispitano samo u nepušača u prva dva izvora: u prvom izvoru prevalencija alergijskih dišnih bolesti bila je malo viša među pasivnim pušačima (23 % naspram 9 %), dok je u drugom izvoru bilo obrnuto (28 % kod pasivnih pušača naspram 35 % kod ostalih), ali povezanosti nisu bile značajne.

Povezanost alergijskog rinitisa i pušenja, aktivnog i pasivnog, ispitana je u nedavnoj meta-analizi autorice Saulyte i suradnika (129). Za obje vrste izloženosti duhanskom dimu bilo je uključenih istraživanja kod kojih je dobiven rizičan učinak, onih kod kojih je učinak bio protektivan, kao i onih u kojima nije dobiven značajan učinak. Zaključak autora meta-analize bio je da aktivno pušenje nije značajno povezano s alergijskim rinitisom: procijenjeni relativni rizik bio je 1,02 s 95 %-tnim intervalom pouzdanosti [0,92 – 1,15]. Ipak, kad se gledao učinak samo u odraslih, dobivena je negativna povezanost (relativni rizik 0,90 [0,82 — 0,99]). Pasivno je pak pušenje bilo značajno pozitivno povezano s alergijskim rinitisom, i kod odraslih (relativni rizik 1,17 [1,03 – 1,32]) i kod djece i adolescenata (relativni rizik 1,09 [1,04 — 1,14]). Ipak, s obzirom da je većina uključenih istraživanja bila presječna, dobivene povezanosti nije moguće interpretirati u uzročno-posljedičnom smislu.

Povezanost izloženosti duhanskom dimu i dišnih bolesti u djece razmatrani su u dva nedavna pregledna istraživanja. Feleszko i suradnici su u svojoj meta-analizi zaključili da je izloženost duhanskom dimu u djece značajno povezana s alergijskom senzibilizacijom te je rizični čimbenik za razvoj alergijskih bolesti u djetinjstvu (130). Burke i suradnici su ranijom meta-analizom 79 prospektivnih istraživanja zaključili da je izloženost pasivnom pušenju u djece zaslužna za povećanje incidencije zviždanja u prsima i astme od 20 % kod djece i mladih (131).

Povezanost izloženosti duhanskom dimu i astme u odraslih opisana je u nedavnim preglednim istraživanjima autora Polose i Thomsona (132) te Toskale i Kennedyja (133). Osim što je istaknuta viša prevalencija astme kod pušača, istaknut je i negativan utjecaj izloženosti duhanskom dimu na zdravlje astmatičara. Naime, kako

navode Polosa i Thomson, aktivno je pušenje kod astmatičara povezano s razvojem težih simptoma, pogoršanjem kvalitete života, češćim egzacerbacijama, većim brojem napadaja astme opasnih po život i općenito višim mortalitetom. Također, terapija kortikosteroidima manje je učinkovita kod astmatičara koji puše. Navodi se također i da je prestanak pušenja povezan s poboljšanjem kliničke slike astme i plućne funkcije. Slično, i pasivno je pušenje povezano s lošijom kontrolom astme, lošijom kvalitetom života i lošijom plućnom funkcijom (133).

Pogoršanje simptoma i kvalitete života uzrokovano pušenjem moglo bi motivirati ispitanike da, nakon pojave bolesti, odustanu od pušenja. To bi mogao biti razlog naizgled protektivnom učinku pušenja u presječnim istraživanjima, kakav je dobiven u meta-analizi pušenja i alergijskog rinitisa u odraslih te u ovom istraživanju.Takav je trend vidljiv i kod pasivnog pušenja u drugom izvoru podataka (administrativni radnici), gdje je prevalencija alergijskih bolesti bila viša kod onih koji nisu pasivno pušili. Jedini suprotan trend bio je vidljiv u prvom izvoru podataka (studenti preddiplomskih studija), gdje je viša prevalencija bolesti bila vidljiva kod pasivnih pušača. Ipak, s obzirom da su to mladi ljudi, vjerojatno je objašnjenje da oni teško mogu izbjeći pasivno pušenje čak i da žele s obzirom da se na mjestima za izlazak uglavnom puši. To se vidi i u vrlo visokoj prevalenciji ispitanika koji su pasivno pušili u ovom izvoru podataka (96 %).

6.2.4.4 Tjelesna aktivnost

Povezanost tjelesne aktivnosti i astme bila je tema meta-analize autorice Eijkemans i suradnika (134). Analiziravši rezultate iz pet tada dostupnih longitudinalnih istraživanja, autori su zaključili da je moguće da tjelesna aktivnost smanjuje rizik za astmu. Naime, procijenjeni omjer šansi iz ovih pet istraživanja bio je 0,88, iako nije bio statistički značajan (95 %-tni interval pouzdanosti [0,77 – 1,01]). Kao objašnjenje za moguću negativnu povezanost navedene su hipoteze o povoljnom djelovanju

tjelesne aktivnosti na dišni sustav, kao što je protuupalno djelovanje (135). Također, kao interpretacija negativne povezanosti u presječnim istraživanjima nekad se navodi da tjelesna aktivnost može potaknuti simptome astme (136) zbog čega bi astmatičari mogli biti skloni izbjegavanju tjelesne aktivnosti.

Rezultati presječnih istraživanja koji su tjelesnu aktivnost procjenjivali upitnikom u meta-analizi Eijkemans i suradnika samo su opisani, bez kvantitativne analize, zbog značajnih razlika među populacijama i korištenim metodama uključenih istraživanja. Trinaest istraživanja pokazalo je značajnu negativnu povezanost tjelesne aktivnosti i astme, tri su pokazala značajnu pozitivnu povezanost, dok u osamnaest istraživanja povezanost tjelesne aktivnosti i astme nije bila statistički značajna.

Učincima tjelesne aktivnosti u astmatičara bavio se Cochraneov pregledni rad iz 2013. (137). Autori su zaključili da bavljenje tjelesnom aktivnošću u astmatičara poboljšava funkciju srca i pluća, a moguće i kvalitetu života. Naime, od pet uključenih istraživanja koji su procjenjivali napredak u kvaliteti života uslijed bavljenja tjelesnom aktivnošću, u četiri je istraživanja dobiven značajan (i statistički i klinički) napredak. Takvi bi učinci mogli motivirati ispitanike s bolesti da se češće bave tjelesnom aktivnošću.

U ovom istraživanju povezanost tjelesne aktivnosti i alergijskih dišnih bolesti ispitana je u dva izvora. U oba izvora ispitanici koji su se dovoljno bavili tjelesnom aktivnošću imali su više šanse za alergijske dišne bolesti. U prvom izvoru podataka (studenti preddiplomskih studija) taj omjer nije bio značajan (1,07 [0,38 - 3,05]), dok je u trećem izvoru (registar prethodnih pregleda u farmaceutskoj industriji) bio uočljivo viši i statistički značajan (3,81 [1,09 - 13,25]). Iako je moguća interpretacija ovakvog rezultata želja za zdravijim životnim stilom i boljom kvalitetom života u ispitanika s bolesti, podaci ne dopuštaju pouzdanije zaključke. Naime, osim što je broj ispitanika s bolesti u ovom izvoru bio dosta nizak (21, 7 %), tjelesna aktivnost ocijenjena je na temelju nedovoljno preciznog pitanja („Bavite li se redovito sportom

ili teško fizički radite i van rada (u slobodno vrijeme)“).

6.2.4.5 Povezanost depresije i alergijskih dišnih bolesti

Povezanost alergijskih dišnih bolesti i depresije predložena je kao moguća u zajedničkom modelu sva četiri izvora podataka. Ispitati je bilo moguće samo povezanost između depresije i alergijske astme u četvrtom izvoru podataka (WHS), a rezultati su naznačili da ispitanici sa simptomima depresije doista imaju veće šanse za astmu, iako razlika nije bila statistički značajna (omjer šansi 2,52 [0,93 – 6,85]).

Povezanost između astme i depresije dobro je poznata u literaturi (138), iako smjer i mehanizam ove povezanosti još uvijek nisu poznati. U nedavnoj meta-analiza velikih prospektivnih istraživanja Gao i suradnici (139) zaključili su da depresija povećava rizik za pojavu astme za 43 % (relativni rizik 1,43 s 95 %-tnim intervalom pouzdanosti [1,28 – 1,61]). Utjecaj astme na pojavu depresije bio je manje jasan: iako je dobivena pozitivna povezanost, nije bila značajna (relativni rizik 1,23 [0,72 – 2,10]), no pouzdanije zaključke nije bilo moguće donijeti budući da su samo dva istraživanja bila uključena u ovu analizu. Pozitivna povezanost između depresije i astme dobivena je i u drugoj nedavnoj meta-analizi. Ispitanici s depresijom imali su oko 3 puta veće šanse za astmu (omjer šansi 3,17 [2,82 – 3,56]), dok su ispitanici s astmom imali oko 1,5 puta veće šanse za depresiju (omjer šansi 1,52 [130 – 1,79]) (140).

Upalni procesi povezani s alergijama navedeni su kao jedan mogući mehanizam koji bi povezivao depresiju i astmu. Prema rezultatima meta-analize autora Jianga i suradnika (140), ispitanici s depresijom imali su značajno više razine citokina IL-4, IL-6 i TNF-alfa u odnosu na ispitanike bez depresije, što bi, s obzirom da ovi citokini igraju ulogu u alergijskim bolestima, išlo u prilog navedenoj hipotezi. Kao drugi mehanizmi koji bi mogli biti poveznica između depresije i astme spominju se i neuroendokrini učinci depresije ili oksidativni stres (139). Naposljetku, nisu

isključeni ni zajednički čimbenici koji bi prethodili obje bolesti, kao što su genetički čimbenici ili izloženost stresu (138).

Povezanost između depresije i drugih atopijskih bolesti istraživana je u manjoj mjeri. Ipak, Chida i suradnici su u meta-analizi prospektivnih istraživanja, koja je uključila prvenstveno radove vezane za astmu, a u manjoj mjeri alergijski rinitis, dermatitis i alergije na hranu, zaključili da su atopijske bolesti povezane s pogoršanjem mentalnog zdravlja, ali i da loše mentalno zdravlje povećava rizik za pojavu atopijskih bolesti (141). Povezanost između depresije i drugih atopijskih bolesti (alergijskog rinitisa i dermatitisa) dobivena je i u drugim nedavnim istraživanjima (142, 143).

6.2.5 Kontakt s mačkom

Od svih ispitivanih životinja (pas, mačka, ptica, glodavac) jedino je kontakt s mačkom bio značajno povezan s alergijskim dišnim bolestima. Povezanost je ispitana samo za stanovnike grada kako bi se isključio mogući utjecaj seoskog okruženja na ovu vezu.

Rezultati o povezanosti kontakta s mačkom i prisutnosti alergijske dišne bolesti nisu bili konzistentni. Dok je u prvom izvoru podataka (studenti preddiplomskih studija) prevalencija bolesti bila značajno viša u ispitanika u kontaktu s mačkom (30 % naspram 20 % kod ispitanika koji nisu bili u kontaktu s mačkom), u trećem je izvoru podataka (registar prethodnih pregleda u farmaceutskoj industriji) bilo obrnuto: nijedan ispitanik koji je držao mačku kao kućnog ljubimca nije imao bolest, u usporedbi s 18 % ispitanika koji nisu imali mačku za ljubimca. U drugom pak izvoru podataka (administrativni radnici) nije dobiven značajna razlika, iako su ispitanici u kontaktu s mačkom imali nešto višu prevalenciju bolesti (32 % naspram 28 %).

Nedosljedni rezultati o povezanosti kontakta s mačkom i alergijskih bolesti do-

biveni su i u ranijim istraživanjima, kako je sažeto u preglednim radovima (144, 145). Kako se navodi u radu autora Konradsena i suradnika (145), teško je utvrditi pravu prirodu povezanosti kontakta s mačkom i razvoja alergija jer je moguće da se ispitanici s alergijskim bolestima, pogotovo oni s težim simptomima, odlučuju na život bez ljubimaca.

6.2.6 Centralno grijanje i alergijske dišne bolesti

Povezanost vrste grijanja u kućanstvu i alergijskih dišnih bolesti još uvijek nije razjašnjena. S jedne strane, poznato je da centralno grijanje isušuje zrak, što može isušiti i nosnu sluznicu i učiniti je pogodnijom za nastanak upala. S druge strane, vlažnija su okruženja plodno tlo za razvoj prašinskih grinja, gljivica i plijesni (146) s posljedičnim povećanjem razina alergena u kućnom okolišu, te je iz tog razloga moguć i povoljan utjecaj centralnog grijanja koje smanjuje vlažnost u stambenom objektu. Primjerice, u velikom europskom istraživanju *European Community Respiratory Health Survey* (ECRHS) vidljive plijesni bile su značajno rjeđe u domovima s centralnim grijanjem (18,9 %) u usporedbi s domovima s ostalim vrstama grijanja (26,8 %; omjer šansi 0,72 [0,64 – 0,81]) (147). Također, postoje preporuke da se za smanjenje količine prašinskih grinja u domu relativna vlažnost zraka održava ispod 50 % (148, 149). Epidemiološka istraživanja o povezanosti količine prašinskih grinja i relativne vlažnosti u domu dala su, međutim, kontradiktorne rezultate (150), a kontradiktorni rezultati dobiveni su i za povezanost koncentracije prašinskih grinja te koncentracije gljivica *Cladosporium* u domu i vrste grijanja (150, 151, 152). Također, u nedavnom Cochraneovom preglednom radu nije pokazan bitan klinički efekt korištenja mehaničkih naprava za smanjivanje vlažnosti kod pacijenata s astmom (146).

Kontradiktorni rezultati dobiveni su i za povezanost astme i rinitisa sa centralnim grijanjem. Primjerice, dok je u turskom istraživanju sa skoro 3000 ispitanika centralno grijanje bilo povezano s povećanim rizikom za alergijske bolesti (omjer šansi 1,26

[1,08 – 1,50]) (153), u ECRHS istraživanju centralno grijanje bilo je povezano samo sa smanjenom učestalošću zviždanja (engl. *wheezing*) i osjećaja nedostatka zraka (omjer šansi 0,87 [0,76 – 0,99]), no ne i s drugim simptomima astme (napadaji gušenja i buđenje noću zbog astmatskog napadaja) i uzimanjem lijekova za astmu (omjer šansi 0,99 [0,83 – 1,18]) (147).

Povezanost centralnog grijanja i alergijskih dišnih bolesti nije bila značajna niti u ovom istraživanju. Naime, iako su u prvom izvoru podataka (studenti) alergijske dišne bolesti bile nešto češće u ispitanika s centralnim grijanjem (24 %) nego u ispitanika s drugim vrstama grijanja (17 %), dok je u drugom izvoru podataka (administrativni radnici) ovaj odnos bio obrnut (prevalencije alergijskih dišnih bolesti od 27 % u ispitanika s centralnim grijanjem i 33 % u ispitanika s drugim vrstama grijanja), ove razlike nisu bile statistički značajne.

Kao i u literaturi, ovakav rezultat moguća je posljedica nepostojanja povezanosti između vrste grijanja i alergijskih bolesti, ali i posljedica interakcije relativne vlažnosti i vrste grijanja s drugim mikroklimatskim uvjetima i navikama ispitanika koje utječu na količinu alergena u domu (147), a nisu bili uključeni u ovo istraživanje (npr. redovito prozračivanje i pranje posteljine, vrsta i starost tepiha, madraca i jastuka, kvaliteta izolacije i slično).

6.2.7 Povezanost socioekonomskog položaja i alergijskih dišnih bolesti

Dok se alergijski rinitis u literaturi povezivao prvenstveno s višim socioekonomskim statusom (2), povezanost astme sa socioekonomskim statusom manje je jasna. Iako je prije i astma bila povezivana s višim socioekonomskim slojem, u posljednje vrijeme sve je više istraživanja koja ne pokazuju povezanost astme i socioekonomskog statusa ili pak pokazuju obrnutu povezanost (viša prevalencija kod ispitanika nižeg

socioekonomskog statusa) (154).

Ovakvi su zaključci dobiveni i u nedavnoj meta-analizi autorice Uphoff i suradnika (155) koja je uključila 183 prethodno objavljena članka, pri čemu su kao pokazatelji socioekonomskog statusa korištene mjere poput obrazovanja ispitanika ili njihovih roditelja, prihoda, veličine kućanstva i dr. Naime, u ovoj su meta-analizi ispitanici nižeg socioekonomskog statusa imali veće šanse za astmu (omjer šansi iz jednostruke regresije 1,38 [1,37 – 1,39]), a manje šanse za alergijski rinitis (omjer šansi 0,52 [0,46 – 0,59]). Slično, za ishod koji je obuhvaćao bilo kakve alergijske bolesti ili atopiju, ispitanici nižeg socioekonomskog statusa imali su manje šanse za bolest (omjer šansi 0,67 [0,62 – 0,72]).

U ovu disertaciju uključene su dvije od navedenih mjera socioekonomskog statusa, stupanj obrazovanja i veličina kućanstva (broj soba po osobi). Mjere su bile povezane: tako su u jedinom izvoru podataka gdje su ove dvije varijable mjerene zajedno (administrativni radnici) ispitanici sa srednjom ili nižom stručnom spremom imali skoro dvaput veće šanse za život u prenapučenim uvjetima, s manje od jedne sobe po osobi, nego ispitanici višeg obrazovanja (omjer šansi 1,92 [1,02 – 3,62]).

Nešto više prevalencije alergijskih dišnih bolesti u ispitanika višeg socioekonomskog statusa uočene su i u ovom istraživanju. Naime, prevalencija bolesti bila je 31 % u ispitanika s barem višom stručnom spremom u drugom izvoru podataka (administrativni radnici), u usporedbi s 25 % kod ispitanika niže stručne spreme. Isti je trend bio prisutan i u trećem izvoru podataka (prethodni pregledi u farmaceutskoj industriji), gdje je 11 % ispitanika s višom stručnom spremom imalo bolest naspram 5 % ispitanika s nižom stručnom spremom. Slično, u drugom izvoru podataka prevalencija bolesti bila je 26 % kod ispitanika koji su živjeli u prenapučenim uvjetima, u usporedbi s 29 % kod ostalih ispitanika. Niža prevalencija bolesti u ispitanika koji su živjeli u prenapučenim uvjetima (19 %) naspram ostalih (24 %) dobivena je i u prvom izvoru podataka (studenti). Ipak, ove razlike nisu bile statistički značajne.

6.2.8 Povezanost neuroticizma i alergijskih dišnih bolesti

Neuroticizam se često opisuje kao osobina ličnosti koja se odnosi na „individualne razlike u negativnom emocionalnom odgovoru na prijetnju, frustraciju ili gubitak" (156) (prema (157, 158)), dok se za osobe s visokim stupnjem neuroticizma često kaže da su „iracionalne" ili pak da imaju „povećanu opću sklonost prema negativnim emocionalnim reakcijama i uznemirenosti" (159). Neuroticizam je povezan sa širokim spektrom mentalnih (npr. poremećaji raspoloženja, tjeskobe, shizofrenija, depresija) i tjelesnih (krvožilne bolesti, sindrom iritabilnog crijeva) poremećaja i oboljenja, većim morbiditetom i mortalitetom bolesti, kao i rizičnim zdravstvenim ponašanjem (156). Ipak, istraživanja o povezanosti neuroticizma i alergijskih dišnih bolesti rijetka su, provedena uglavnom u malim uzorcima (manje od 200 ispitanika) i s kontradiktornim rezultatima. Pritom se o smjeru i mehanizmu eventualne povezanosti još uvijek samo nagađa. S jedne strane, pretpostavlja se da bi osobe s visokim stupnjem neuroticizma zbog svoje negativne percepcije mogle biti svjesnije simptoma i pretjerivati u njihovu prijavljivanju (160, 161), ili pak da bi utjecaj na alergijske bolesti mogao biti posredovan pretjeranim doživljajem stresa kod neurotičnih ispitanika (162). S druge strane, pretpostavlja se i da bi alergijske bolesti mogle doprinijeti negativnom doživljaju te na taj način dovesti do većeg izražaja neuroticizma u bolesnika (161).

Malo je istraživanja koja su u velikom uzorku (od barem 1000 ispitanika) ispitivala povezanost neuroticizma i alergijskih dišnih bolesti. U nedavnom presječnom istraživanju autorice Cheng i suradnika s oko 6000 ispitanika neuroticizam je bio statistički značajno povezan s alergijskim rinitisom (omjer šansi 0,92 [0,85 – 0,99] za niže u usporedbi s višim vrijednostima neuroticizma) (163). Pozitivna veza dobivena je i u presječnom istraživanju autora Goodwina i suradnika s oko 3000 ispitanika, gdje je neuroticizam bio značajno povezan s alergijama u ukupnom uzorku (omjer šansi 1,22 [1,04 – 1,43]) u muškaraca (omjer šansi 1,43 [1,13 – 1,82]), ali ne i u žena

(omjer šansi 1,07 [0,87 – 1,33]) (161).

Tri su prospektivna istraživanja u velikom uzorku ispitivala povezanost neuroticizma i astme, ali bez razlikovanja alergijske i nealergijske astme. U istraživanju autorice Huovinen i suradnika (164) s oko 11000 ispitanika povezanost neuroticizma i astme dobivena je samo u presječnom dijelu istraživanja u ukupnom uzorku koji je uključivao oba spola, uz kontrolu za dob i spol (omjer šansi 1,78 [1,12 – 2,84]), dok u višestrukoj regresiji s ostalim bitnim čimbenicima, provedenoj zasebno u muškaraca i žena, učinak neuroticizma nije bio značajan niti u jednog spola. Također, nije bilo niti značajnog učinka neuroticizma na budući nastanak astme (omjeri šansi iz višestrukih regresija 1,36 [0,64 – 2,87] kod muškaraca i 0,60 [0,29 – 1,21] kod žena) u prospektivnom dijelu istraživanja. S druge strane, u druga dva prospektivna istraživanja neuroticizam je bio značajno povezan s nastankom astme u periodu praćenja. Tako je u istraživanju autora Loerbroksa i suradnika s oko 4000 ispitanika (162) učinak neuroticizma bio prisutan u cjelokupnom uzorku (relativni rizik za nastanak astme 3,07 [1,71 – 5,48]), ali i zasebno kod muškaraca i žena, kao i u drugom istraživanju Loerbroksa i suradnika (165) s oko 12000 ispitanika (relativni rizik 1,59 [1,12 – 2,25] u cjelokupnom uzorku, bez posebne analize po spolovima).

Ostala istraživanja s manjim uzorcima, uključujući i prethodno istraživanje s podacima iz drugog izvora podataka (166), dobila su kontradiktorne rezultate o presječnoj povezanosti neuroticizma s astmom i alergijskim rinitisom (160, 166, 167) te o povezanosti neuroticizma i pridržavanja terapije za ove bolesti (168, 169).

Povezanost neuroticizma i alergijskih dišnih bolesti nije dobivena niti u ovom istraživanju. Kao i u prethodno objavljenom radu (166), povezanost nije bila značajna u drugom izvoru podataka (administrativni radnici), iako je prevalencija bolesti bila nešto viša u ispitanika s višim stupnjem neuroticizma (31 % naspram 27 % u ispitanika s nižim stupnjem neuroiticizma). U prvom izvoru podataka (studenti) ispitanici s višim stupnjem neuroticizma imali su nešto nižu prevalenciju bolesti (21

%) nego oni s nižim stupnjem neuroticizma (23 %), međutim, ni ova razlika nije bila statistički značajna.

6.2.9 Povezanost prehrane s alergijskim dišnim bolestima

U literaturi je predložena hipoteza da bi se porast prevalencije atopijskih bolesti diljem svijeta mogao dijelom pripisati širenju takozvanog zapadnjačkog načina prehrane, koji karakterizira povećan unos obrađene hrane i omega-6 masnih kiselina te smanjeni unos antioksidansa, prvenstveno putem voća i povrća, i omega-3 masnih kiselina (154). U ovom je istraživanju ispitivan samo unos voća i povrća, pri čemu je kategorizacija unosa voća i povrća provedena prema preporukama Svjetske zdravstvene organizacije za prevenciju kroničnih bolesti (51, 52).

Istraživanja o povezanosti unosa voća i povrća i alergijskih dišnih bolesti pretežno su opažajna i velikim se dijelom odnose na astmu. Dvije nedavne meta-analize o povezanosti astme i unosa voća i povrća ukazale su na zaštitni učinak voća i povrća. Meta-analiza autora Nurmatova i suradnika (170) uključila je 22 istraživanja o unosu voća i povrća, od kojih je 17 pokazalo pozitivan učinak voća i povrća. Ipak, zbog velikih razlika u dizajnu istraživanja, dobi ispitanika i mjerenim ishodima, sumarni omjer šansi izračunat je iz samo četiri istraživanja za unos voća (omjer šansi za zviždanje u prsima 0,75 [0,60 – 0,94]) te iz tri istraživanja za unos povrća (omjer šansi za zviždanje 0,84 [0,64 – 1,10]), pri čemu su ispitanici bili djeca u dobi od 10 do 14 godina. Iako je sama meta-analiza ocijenjena kao kvalitetna, jačina dokaza ocijenjena je kao „vrlo niska" jer je sumarni učinak dobiven iz malog broja opažajnih istraživanja s visokom heterogenosti rezultata (171).

U novijoj meta-analiza autora Seyedrezazadeh i suradnika (172) viši unos voća bio je povezan s manjim izgledima za astmu u odraslih i djece (omjer šansi 0,84 [0,79 – 0,89]), kao i viši unos povrća (omjer šansi 0,88 [0,82 – 0,95]). Međutim, i u ovoj meta-analizi većina istraživanja bila je opažajna, a heterogenost rezultata velika, te

je istaknuta potreba za više istraživanja praćenja i kontroliranih pokusa kako bi se dobili pouzdaniji rezultati.

Opažajna istraživanja o povezanosti unosa antioksidansa i alergijskih bolesti kritizirana su i zbog načina mjerenja unosa i korištenih statističkih metoda. Naime, kako se detaljno opisuje u radu autora Allana i suradnika (173), većina ovih istraživanja kao metodu mjere unosa namirnica koristi prehrambeni upitnik (engl. *food frequency questionnaire*, FFQ), pri čemu neka istraživanja uključuju samo grube kategorije malog dijela namirnica, a neka istraživanja detaljna pitanja o velikom broju namirnica, što dovodi do velikih razlika u kvaliteti procjene unosa. FFQ je kritiziran i zbog pristranosti u prisjećanju ispitanika te iskrivljivanju vlastitog unosa kako bi se odao dojam zdravijeg životnog stila te se općenito smatra metodom niske točnosti. Točnije mjere unosa antioksidansa, poput analize serumske krvi, skupe su i uglavnom su korištene na način da odražavaju samo nedavan unos hrane, no ne i dugotrajan učinak prehrane. Naposljetku, ova istraživanja često ne korigiraju unos antioksidansa za cjelokupni energetski unos ispitanika, kao ni za višestruko statističko testiranje, što dovodi do iskrivljene slike učinka (vidi (173)).

Većina spomenutih problema odnosi se i na podatke o unosu voća i povrća i u ovom istraživanju. U prvom izvoru podataka (studenti) unos voća i povrća bio je podjednak i kod ispitanika s alergijskom dišnom bolesti (58 % ispitanika s nedovoljnim unosom) i kod ispitanika bez bolesti (60 %) te nije bilo statistički značajne razlike.

6.2.10 Dob i alergijske dišne bolesti

Analizirajući podatke izvora WHS u odrasloj populaciji Hrvatske prevalencija alergijske astme, definirane na način opisan u Metodama (poglavlje 4.1), iznosila je 2003. godine 5,1 % (4,9 % u žena, 5,5 % u muškaraca). Slična prevalencija alergijske astme nađena je i u populaciji hrvatskih adolescenata (5,6 % u Splitsko-dalmatinskoj žu-

paniji) (174).

Prevalencija alergijske astme bila je statistički značajno povezana s dobi u četvrtom izvoru podataka (WHS), međutim, povećanje prevalencije s dobi bilo je minimalno (omjer šansi 1,02 [1,01 - 1,04]). Povezanost nije nađena u drugom izvoru podataka (administrativni radnici), no potrebno je naglasiti da je u ovoj bazi udio ispitanika mlađih od 30 godina samo 7 % (11 ispitanika), a udio ispitanika starijih od 59 godina svega 3 % (5 ispitanika). Prvi izvor podataka bio je dobno homogen (studenti prve godine fakulteta s medijanom dobi od 19 godina), dok je prevalencija astme u trećem izvoru podataka (prethodni pregledi u farmaceutskoj industriji) bila premala (< 1 %) za analizu.

U izvorima analiziranim u ovom istraživanju analizirana je samo dobna prevalencija astme, dok se incidencija astme nije mogla izračunati s obzirom da dob nastanka astme nije bilo moguće utvrditi sa sigurnošću. Poznato je da alergijska senzibilizacija opada s dobi (175), pa tako i incidencija alergijske astme (98). Povezanost dobi s prevalencijom astme, međutim, nije toliko jasna s obzirom da prevalenciji astme u odrasloj dobi doprinose kako slučajevi astme koji perzistiraju od djetinjstva i egzacerbacije astme nastale još u djetinjstvu, tako i novonastali slučajevi astme u odrasloj dobi (engl. *adult-onset asthma*) (176). Naime, iako je astma koja nastaje u djetinjstvu u većoj mjeri povezana s atopijom i muškim spolom, a astma koja nastaje u odraslih u većoj se mjeri povezuje sa ženskim spolom i okolišnim čimbenicima (npr. pušenje ili izloženost inhalatornim štetnostima u radnom okolišu), alergijska senzibilizacija ima ulogu i u dijelu slučajeva astme koja nastaje u odraslih (176).

Istraživanja alergijskog rinitisa pokazuju trend smanjenja prevalencije s dobi iznad 45 godina života (79, 177, 178, 179). U izvorima uključenim u ovo istraživanje nije bilo statistički značajnog utjecaja dobi u drugom i trećem izvoru podataka. Međutim, zbog malog broja slučajeva u pojedinim dobnim skupinama (manje od 15 slučajeva bolesti u gotovo svim dobnim skupinama s rasponom od 10 godina, te

manje od 5 slučajeva u polovici dobnih skupina), nije moguće donositi pouzdanije zaključke o trendu.

6.2.11 Povezanost mjesta stanovanja i alergijskih dišnih bolesti

Život u urbanim okruženjima u literaturi se povezuje s višom prevalencijom alergijskih dišnih bolesti u odnosu na ruralno okruženje (154, 180, 181). Ovi se učinci povezuju s onečišćenjem zraka, prvenstveno u područjima s gustim prometom, ali i sa zaštitnim učinkom ruralnog okruženja, pogotovo u djetinjstvu („higijenska teorija") (182).

Kao čestice i spojevi koji se mogu naći u onečišćenom zraku, a dovedeni su u vezu s dišnim bolestima, izdvojene su čestice $PM_{2.5}$ i PM_{10}, dušikovi oksidi (naročito dušikov monoksid i dioksid), ugljični monoksid, sumporov dioksid i ozon. U posljednje je vrijeme provedeno nekoliko meta-analiza o povezanosti ovih spojeva s dišnim bolestima, i to prvenstveno s astmom.

Od dušikovih oksida, povezanost dušikova dioksida s astmom pronađena je u meta-analizama autora Gasane i suradnika (omjer šansi 1,14 [1,06 – 1,24] za incidenciju astme, za povećanje izloženosti od 10 μg NO_2 po metru kubnom) (183) te u meta-analizi autora Andersona i suradnika (omjer šansi 1,12 [1,04 – 1,21)]) (184), dok u meta-analizi autora Bowattea i suradnika nije bilo povezanosti (omjer šansi 1,09 [0,96 – 1,23]) (185). U meta-analizi autora Gasane i suradnika i izloženost ugljičnom monoksidu bila je povezana s većim izgledima za astmu u djetinjstvu (omjer šansi 1,06 [1,01 – 1,12]), dok povezanost izloženosti ozonu i astme nije pronađena (omjer šansi 1,00 [0,94 – 1,08]) (183). U istoj su meta-analizi izloženosti sumporovom dioksidu i PM česticama bile povezane sa simptomom zviždanja u prsima (omjeri šansi 1,04 [1,01 – 1,07] i 1,05 [1,04 – 1,07]), dok izloženost sumporovom dioksidu nije bila

povezana s astmom (omjer šansi 1,02 [0,96 – 1,08]). Također, nije pronađena ni povezanost $PM_{2.5}$ čestica i astme (omjer šansi 1,06 [0,93 – 1,21] za prevalenciju te 1,40 [0,77 – 2,56] za incidenciju astme), kao ni u meta-analizama autora Andersona i suradnika (omjer šansi 1,16 [0,98 – 1,37] za incidenciju astme ili simptoma zviždanja) (184) i Bowattea i suradnika (omjer šansi 1,14 [1,00 – 1,30] za incidenciju astme, pri povećanju izloženosti od 2 μg po metru kubnom) (185) te se učinci onečišćenja još uvijek smatraju malima (186).

Povezanost onečišćenja zraka s alergijskom senzibilizacijom i alergijskim rinitisom ispitivana je u manjem broju istraživanja. Snaga dokaza slabija je nego u slučaju astme te postoji potreba za bolje dizajniranim istraživanjima (187). Tako u meta-analizi autorice Gruzieve i suradnika nije dobivena povezanost onečišćenja zraka s alergijskom senzibilizacijom (188), dok su za alergijski rinitis dobiveni kontradiktorni rezultati u analizi autora Bowattea i suradnika (185). Naime, od pet istraživanja uključenih u ovaj pregledni rad, samo su dva pokazala povezanost onečišćenja zraka s alergijskim rinitisom.

Zaštitni učinak ruralnog okruženja, naročito seoskih gospodarstava (farmi) u kojima je prisutan uzgoj domaćih životinja i biljna proizvodnja, prepoznat je u literaturi (180), a dovodi se u vezu s pojačanom izloženosti imunološkog sustava bakterijama, parazitima i gljivicama te različitim alergenima biljnog i životinjskog porijekla u ranom djetinjstvu, u doba sazrijevanja imunološkog sustava.

U ovom istraživanju prevalencija alergijskih dišnih bolesti bila je u sva tri izvora podataka (studenti, administrativni radnici i prethodni pregledi u farmaceutskoj industriji) nešto viša u ispitanika s trenutnim boravištem u gradu (prevalencije bolesti redom 23 %, 29 % i 8 % u ispitanika iz grada, naspram 14 %, 19 % i 3 % u ispitanika sa sela), no razlika nije bila statistički značajna. Nažalost, nije bilo moguće mjeriti izloženosti spomenutim spojevima, što bi dalo jasniju sliku o povezanosti s onečišćenjem zraka, a niti utvrditi mjesto prebivanja u ranom djetinjstvu

(urbano ili ruralno) za veliku većinu ispitanika, kako bi se procijenila uloga higijenske teorije u izvorima uključenim u ovo istraživanje.

6.2.12 Povezanost kvalitete života i alergijskih dišnih bolesti

Poznato je da bilo koja kronična bolest, pa tako i alergijske bolesti dišnog sustava, negativno utječu na kvalitetu života, kako u djece tako i u odraslih (189, 190, 191). Opisano je, primjerice, kako alergijski rinitis smanjuje kvalitetu spavanja (uslijed simptoma bolesti koje remete normalan ciklus spavanja), ograničava dnevne aktivnosti, te negativno utječe na fizički i mentalni status i socijalno funkcioniranje oboljelih osoba (191). Pritom se u epidemiološkim istraživanjima kvaliteta života istražuje pomoću subjektivne procjene brojnih aspekata, uključujući materijalno, socijalno, emocionalno i tjelesno stanje osobe (192).

Kvaliteta života u ovom je istraživanju bila predstavljena pitanjem „Koliko ste zadovoljni svojim zdravljem", s mogućim odgovorima na skali od 1 („vrlo nezadovoljan") do 5 („vrlo zadovoljan"). Navedeno pitanje jedno je od dva osnovna pitanja koja se u kratkom upitniku o kvaliteti života Svjetske zdravstvene organizacije (WHOQOL-BREF) koriste u subjektivnoj procjeni općeg zdravstvenog stanja i kvalitete života osobe (54). Ovo je pitanje odabrano kao pitanje iz domene kvalitete života prisutno u dva izvora podataka (administrativni radnici - drugi izvor podataka i WHS - četvrti izvor podataka).

Prosječna ocjena zadovoljstva zdravljem na skali od 1 do 5 u ispitanika u ovom istraživanju (3,70 u drugom te 3,66 u četvrtom izvoru podataka) bila je vrlo slična prosječnim ocjenama dobivenim u drugim istraživanjima u hrvatskoj populaciji (3,63 u istraživanju autorica Tucak i Nekić (192) te 3,64 i 3,70 (originalni rezultat preračunat na skalu od 1 do 5) u dva istraživanja autorica Kaliterna Lipovčan i Prizmić-Larsen (193)).

Prijašnja su istraživanja pokazala da je zadovoljstvo zdravljem usko povezano

s dobi i socioekonomskim statusom. Tako su mlađi ispitanici te ispitanici viših prihoda, kao i zaposleni ispitanici u odnosu na nezaposlene i umirovljene, bili zadovoljniji zdravljem (192, 194, 195). Ovaj je trend bio prisutan i u ovom istraživanju. Tako su u izvoru WHS zadovoljniji zdravljem bili i mlađi ispitanici i zaposleni, pri čemu je dobni trend bio puno izraženiji kod ispitanika koji nisu bili stalno zaposleni (ocjene zadovoljstva zdravljem redom 4,16, 3,56 i 3,20 u dobnim skupinama 18 – 40, 41 – 60 i više od 60 u nezaposlenih, u usporedbi s ocjenama od 3,94, 3,80 i 3,82 u zaposlenih ispitanika). U drugom izvoru podataka svi su ispitanici bili zaposleni pa su i ocjene manje varirale, iako je bilo vidljiv pad u zadovoljstvu zdravljem kod najstarijih ispitanika (3,70, 3,71 i 3,40 redom u dobnim skupinama 18 – 40, 41 – 60 i više od 60).

U WHS izvoru bilo je moguće ispitati povezanost zadovoljstva zdravljem samo s alergijskom astmom. Pritom su ispitanici s astmom imali nižu prosječnu ocjenu zadovoljstva zdravljem (3,05, što je usporedivo s ocjenom od 2,90 kod ispitanika s kroničnim respiratornim bolestima (195)) te je bilo više nezadovoljnih zdravljem (30 %) nego kod ostalih ispitanika (prosječna ocjena 3,69, 13 %). Ova je razlika bila statistički značajna: ispitanici nezadovoljni zdravljem imali su gotovo tri puta veće šanse za bolest (omjer šansi 2,95 [1,30 – 6,68]).

Povezanost zadovoljstva zdravljem i alergijskih dišnih bolesti (uključujući rinitis) bilo je moguće ispitati samo u drugom izvoru podataka. Iako je među ispitanicima s astmom ili rinitisom i ovdje bilo više nezadovoljnih zdravljem (11 %, prosječna ocjena 3,59) u odnosu na ostale (8 %, prosječna ocjena 3,74), razlika nije bila statistička značajna. Moguće objašnjenje je da se ovdje radilo o zaposlenim, radno sposobnim ljudima, u prosjeku mlađim nego u izvoru WHS, kod kojih nema toliko varijacija u zadovoljstvu zdravljem.

6.2.13 Usporedba rezultata različitih vrsta analize

Za kraj, kako je već opisano u prethodnim dijelovima ove Rasprave, bilo je vidljivo neslaganje različitih vrsta analize (logistička regresija, Bayesova mreža jednog izvora podataka, Bayesova mreža više izvora podataka) glede povezanosti nekih čimbenika s alergijskim dišnim bolestima.

U usporedbi logističke regresije i Bayesove mreža jednog izvora podataka, vidljivo je da je analiza Bayesovom mrežom jednog izvora podataka manji broj čimbenika izravno povezala s alergijskom dišnom bolesti nego analiza logističke regresije (Tablica 24). Tako su muški spol u prvom izvoru (studenti preddiplomskih studija) te alergijske bolesti u roditelja i redovita tjelesna aktivnost u trećem izvoru (registar prethodnih pregleda u farmaceutskoj industriji) bile značajno povezane s alergijskim dišnim bolestima samo u modelu logističke regresije, dok su se sve ostale izravne povezanosti podudarale u ova dva modela.

Budući da se radi o epidemiološkim podacima o čijim stvarnim uzročno- posljedičnim mehanizmima možemo samo nagađati, ne možemo znati kojom su vrstom analize dobiveni rezultati najbliže istini. Jedan mogući razlog za ovakvo neslaganje je što Bayesove mreže, u usporedbi s logističkom regresijom, drugačije modeliraju međuodnos čimbenika pa utjecaji koji su u logističkoj regresiji nađeni kao izravni u modelu mreže postaju neizravni (primjerice, neizravan utjecaj spola na alergijske dišne bolesti putem indeksa tjelesne mase u Bayesovoj mreži prvog izvora). Drugi mogući razlog je što, kako je diskutirano u poglavlju 6.1.2, korišteni algoritam za Bayesove mreže ne uspijeva otkriti sve prisutne povezanosti, što je moguće objašnjenje za neprepoznatu povezanost alergijskih bolesti u roditelja i alergijske dišne bolesti u trećem izvoru podataka.

Povezanosti dobivene analizom jednog izvora podataka koje nisu zadržane u analizi više izvora podataka (primjerice, povezanost aktivnog pušenja i bolesti) mogu se objasniti time što analiza više izvora podataka uzima u obzir da li je povezanost

dobivena u svim izvorima podataka u kojima je ispitivana ili samo u nekim, pri čemu koristi algoritam za rješavanje kontradiktornih rezultata temeljen na kriteriju maksimalnog posteriornog omjera (engl. *maximum posterior ratio*) (40). S druge strane, povezanosti koje su bile prisutne samo u modelu više izvora podataka (povezanost depresije i konzumacije alkohola s bolesti), u ovom su modelu označene kao moguće vjerojatno iz razloga što ove varijable nisu mjerene zajedno ni u jednom izvoru podataka.

6.2.14 Zaključno

U ovom je istraživanju predložena nova metodologija za procjenu parametara u meta-analizi te je primijenjena na problem izgradnje Bayesovih mreža stapanjem podataka iz više izvora. Očekuje se da će ova metodologija biti vrlo korisna za epidemiološka istraživanja bolesti s višeuzročnom etiologijom, poput alergijskih dišnih bolesti, gdje se očekuje interakcija čimbenika od utjecaja za nastanak bolesti.

Standardna meta-analiza također se može gledati kao metoda stapanja podataka iz više izvora. U odnosu na analizu podataka iz jednog istraživanja, prepoznata je korist meta-analize kao metode koja stavlja rezultate pojedinog istraživanja, koja imaju svoje manjkavosti poput veličine uzorka i pristranosti, u kontekst drugih dobivenih rezultata u području i tako daje širu sliku o utjecaju čimbenika na bolest.

Te bismo dobrobiti očekivali i ovdje. U odnosu na meta-analizu koja procjenjuje samostalni učinak jednog čimbenika, dodana vrijednost ovakve vrste analize je procjena učinka više čimbenika istovremeno, s naglaskom na njihov međuodnos.

Kao i svaka primjena sa stvarnim podacima, i primjena metodologije u ovom istraživanju (čimbenici rizika alergijskih dišnih bolesti i njihov međuodnos) ima svoja ograničenja, zbog kojih je dobivene epidemiološke rezultate potrebno interpretirati s oprezom. Kao prvo, u analizu su bila uključena samo četiri izvora podataka, što je malen broj i u običnoj meta-analizi podataka iz iste populacije. Nadalje, pojedinačne

veličine uzorka nisu bile velike (za tri od četiri izvora podataka broj ispitanika bio je manji od 500), zbog čega je moguće da Bayesove mreže iz pojedinačnih izvora nisu pouzdano procijenjene. Posljedično, s obzirom da su pojedinačne mreže proslijeđene kao ulazni podaci algoritmu za stapanje mreža, moguće je da ni zajednički model sva četiri izvora podataka ne odražava realno međuodnose čimbenika. Kako je već naglašeno, u ovom je koraku izrazito bitna pomoć stručnjaka, pogotovo kod malih veličina uzoraka. Naposljetku, s obzirom na heterogenost korištenih izvora prema periodima provođenja istraživanja i sub-populacija uključenih u njima (studenti i radnici sa specifičnim zanimanjima u tri od četiri izvora podataka), teško je reći da li je u ovoj primjeni stapanje bilo primjereno ili je razumnije analizirati svaki izvor podataka zasebno. Evaluacija na testnim skupovima, koja je trebala odgovoriti na ova pitanja, nije dala jasne zaključke. Gledano ukupno prema svim ispitivanim mjerama valjanosti, model sva četiri izvora nije bio lošiji od metoda analize jednog izvora podataka, a bio je najbolji prema osjetljivosti i specifičnosti, zbog čega se ne može reći da stapanje više izvora u ovoj primjeni nije imalo smisla. S druge strane, model nije bio najbolji prema točnosti i prediktivnim vrijednostima, zbog čega se njegovi rezultati ne mogu uzeti ni kao najpouzdaniji od svih. Pravu prirodu međuodnosa između čimbenika rizika u našoj populaciji moglo bi razjasniti stapanje većeg broja izvora podataka s većim veličinama uzorka iz opće populacije, za što će predložena metodologija svakako biti korisna.

7 Zaključak

1. U istraživanju čimbenika rizika povezanih s alergijskim dišnim bolestima, metoda stapanja podataka iz više izvora pokazala je veću osjetljivost i specifičnost, dok je analiza samo jednog izvora podataka imala veće prediktivne vrijednosti. Analiza više izvora podataka pokazala je najveću prednost u odnosu na analizu jednog izvora podataka u slučaju malog uzorka.

2. U izgradnji strukture kauzalnog modela preporučljivo je koristiti znanje stručnjaka iz područja i literaturno znanje u što većoj mjeri. Vanjsko znanje izrazito je bitno u situacijama kad su veličine uzoraka male te povezanosti ispitane statističkim metodama ostaju neprepoznate zbog manjka statističke snage.

3. Kad su jednom određene uzročno-posljedične veze među varijablama, za procjenu njihovih jačina preporučljivo je (ukoliko je to, s obzirom na broj i heterogenost izvora podataka, tehnički izvedivo) koristiti sve dostupne podatke jer na taj način dobivamo točniju procjenu veličine efekta (primjerice, omjera šansi ili rizika za bolest) i intervala pouzdanosti.

4. U pojedinačnim analizama epidemioloških izvora podataka iz hrvatske populacije, čimbenici najsnažnije povezani s prisutnošću alergijske dišne bolesti bili su alergijske bolesti u djetinjstvu te alergijske bolesti u roditelja. Prekomjerna tjelesna težina i pušenje bili su povezani sa značajno manjim, a redovita tjelesna aktivnost sa značajno većim šansama za prisutnost bolesti, moguće kao rezultat želje za zdravijim životnim stilom i boljom kvalitetom života ispitanika s bolesti. Veće šanse za bolest imali su i muškarci, što je vjerojatno posljedica više prevalencije atopije kod muških ispitanika.

5. U usporedbi s analizama pojedinačnih izvora, metoda stapanja podataka iz

više izvora prepoznala je manje izravnih povezanosti s bolesti (alergijske bolesti u djetinjstvu, alergijske bolesti u roditelja, indeks tjelesne mase), dok su ostale bile neizravne (spol, tjelesna aktivnost), ili nisu prepoznate (pušenje). S druge strane, metoda je predložila dvije nove veze (povezanost depresije i konzumacije alkohola s bolesti) između varijabli koje nisu niti u jednom izvoru ispitivane zajedno. Uočene razlike mogu biti odraz stvarnih međuodnosa, nedostatne statističke snage ili karakteristike metode da ne prepoznaje sve prisutne veze.

6. Najveći doprinos opisane metode stapanja više izvora podataka očekuje se u procjeni jačine veza, uz pomoć ekspertnog znanja pri procjeni uzročno-posljedičnih veza. Pouzdanost procjene moguće je povećati uključivanjem većeg broja izvora s većim uzorcima podataka.

8 Kratki sadržaj na hrvatskom jeziku

Pretpostavlja se da bi zajednička analiza više izvora podataka mogla pomoći u razjašnjenju etiologije nezaraznih, kroničnih bolesti. Međutim, statističke metode nisu dovoljno razvijene za situacije kad se različita istraživanja samo djelomično preklapaju prema mjerenim čimbenicima.

U ovom je istraživanju predložena nova metoda za meta-analizu takvih podataka, G-MA, i primijenjena u izradi Bayesovih mreža više izvora podataka. U računalnim eksperimentima G-MA je imala manju srednju kvadratnu pogrešku i uže intervale pouzdanosti u procjeni jačine veza u odnosu na analizu pojedinačnih izvora.

U istraživanju čimbenika povezanih s prisutnošću alergijskih dišnih bolesti Bayesova mreža s G-MA metodom (Bayes-G-MA) primijenjena je na četiri epidemiološka izvora hrvatske populacije. U usporedbi s analizama pojedinačnih izvora, Bayes-G-MA je prepoznala manje izravnih povezanosti s bolesti (alergijske bolesti u djetinjstvu, alergijske bolesti u roditelja, indeks tjelesne mase), dok su ostale bile neizravne (spol, tjelesna aktivnost), ili nisu prepoznate (pušenje). Bayes-G-MA predložila je dvije nove veze (povezanost depresije i konzumacije alkohola s bolesti) između varijabli koje nisu niti u jednom izvoru ispitivane zajedno. Uočene razlike mogu biti odraz stvarnih međuodnosa, nedostatne statističke snage ili karakteristike metode da ne prepoznaje sve prisutne veze.

Zaključno, najveći doprinos metode očekuje se u procjeni jačine veza, uz pomoć ekspertnog znanja pri procjeni uzročno-posljedičnih veza.

9 Kratki sadržaj i naslov doktorskog rada na engleskom jeziku

Exploring the risk factors for the development of allergic respiratory diseases by Bayesian networks built on multiple-source data

It is suggested that multi-source data analysis could help to elucidate the aetiology of non-communicable, chronic diseases. However, statistical methods for multi-source data analysis with partially overlapping data sources are not developed enough.

The study proposes new method for meta-analysis of such data, G–MA, applying it to Bayesian network of multi-source data. In simulation experiments, G-MA showed lower mean squared error and narrower confidence intervals in parameter estimation, when compared to single data source analyses.

Bayesian network with G-MA method (Bayes–G–MA) was applied to analyse the factors related to the presence of allergic respiratory diseases (ARD) in four Croatian epidemiological sources. Compared to single source analyses, Bayes–G–MA recognised lower number of direct relationships with ARD (allergic diseases in childhood, allergic diseases in parents, body mass index), while others became indirect (gender, physical activity) or unrecognised (smoking). Bayes–G–MA suggested two new relationships (with depression and alcohol consumption) between variables not tested together in any single source. These results could reflect true relationships, but also lack of statistical power or intrinsic property of the method to under-report present relationships.

In conclusion, the main contribution of proposed method is in the assessment of

parameters, with expert help needed in evaluation of causal relationships.

Jelena Kovačić, 2017.

10 Popis literature

[1] Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention [Internet]. 2016 [pristupljeno 8.3.2017.]. Dostupno na: http://www.ucalgary.ca/icancontrolasthma/files/icancontrolasthma/wms-gina-2016-main-report-final.pdf.

[2] Bousquet J, Khaltaev N, Cruz AA, Denburg J, Fokkens WJ, Togias A, i sur. Allergic rhinitis and its impact on asthma (ARIA) 2008 update (in collaboration with the World Health Organization, GA(2)LEN and AllerGen). Allergy. 2008;63(Suppl 86):8–160.

[3] Johansson SGO, Bieber T, Dahl R, Friedmann PS, Lanier BQ, Lockey RF, i sur. Revised nomenclature for allergy for global use: Report of the Nomenclature Review Committee of the World Allergy Organization, October 2003. J Allergy Clin Immunol. 2004;113(5):832–6.

[4] Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015;386(9995):743–800.

[5] Lafeuille MH, Gravel J, Figliomeni M, Zhang J, Lefebvre P. Burden of illness of patients with allergic asthma versus non-allergic asthma. J Asthma. 2013;50(8):900–7.

[6] World Allergy Organization. The WAO White Book on Allergy 2013

Update. Pawankar R, Holgate ST, Canonica GW, Lockey RF, Blaiss MS, ur. Milwaukee: World Allergy Organization; 2013.

[7] World Health Organization. Global surveillance, prevention and control of chronic respiratory diseases: A comprehensive approach. Geneva: World Health Organization; 2007.

[8] Zuberbier T, Lötvall J, Simoens S, Subramanian SV, Church MK. Economic burden of inadequate management of allergic diseases in the European Union: a GA2LEN review. Allergy. 2014;69(10):1275–9.

[9] Gupta J, Johansson E, Bernstein JA, Chakraborty R, Khurana Hershey GK, Rothenberg ME, i sur. Resolving the etiology of atopic disorders by using genetic analysis of racial ancestry. J Allergy Clin Immunol. 2016;138(3):676–99.

[10] Harb H, Renz H. Update on epigenetics in allergic disease. J Allergy Clin Immunol. 2015;135(1):15–24.

[11] Bousquet J, Anto JM, Akdis M, Auffray C, Keil T, Momas I, i sur. Paving the way of systems biology and precision medicine in allergic diseases: the MeDALL success story: Mechanisms of the Development of ALLergy; EU FP7-CP-IP; Project No: 261357; 2010-2015. Allergy. 2016;71(11):1513–1525.

[12] Castro-Rodriguez JA, Forno E, Rodriguez-Martinez CE, Celedón JC. Risk and protective factors for childhood asthma: what is the evidence? J Allergy Clin Immunol Pract. 2016;4(6):1111–22.

[13] Tedeschi A, Barcella M, Bo GA, Miadonna A. Onset of allergy and asthma symptoms in extra-European immigrants to Milan, Italy: possible role of environmental factors. Clin Exp Allergy. 2003;33(4):449–54.

[14] Dunlop J, Matsui E, Sharma HP. Allergic rhinitis: environmental determinants. Immunol Allergy Clin North Am. 2016;36(2):367–77.

[15] Schoenwetter WF. Allergic Rhinitis: Epidemiology and Natural History. Allergy Asthma Proc. 2000;21(1):1–6.

[16] Custovic A, Marinho S, Simpson A. Gene-environment interactions in the development of asthma and atopy. Expert Rev Respir Med. 2012;6(3):301–8.

[17] Demchuk E, Yucesoy B, Johnson VJ, Andrew M, Weston A, Germolec DR, i sur. A statistical model for assessing genetic susceptibility as a risk factor in multifactorial diseases: lessons from occupational asthma. Environ Health Perspect. 2007;115(2):231–4.

[18] Choudhry S, Avila PC, Nazario S, Ung N, Kho J, Rodriguez-Santana JR, i sur. CD14 tobacco gene–environment interaction modifies asthma severity and Immunoglobulin E levels in Latinos with asthma. Am J Respir Crit Care Med. 2005;172(2):173–82.

[19] Lynch S, Wood RA, Boushey H, Bacharier LB, Bloomberg GR, Kattan M, i sur. Effects of early-life exposure to allergens and bacteria on recurrent wheeze and atopy in urban children. J Allergy Clin Immunol. 2014;134(3):593–601.

[20] Ramette A, Kuehni CE. Bed-sharing and childhood asthma: from associations to causal modelling. Eur Respir J. 2015;45(3):596–600.

[21] Koller D, Friedman N. Probabilistic Graphical Models. Principles and techniques. Cambridge: The MIT Press; 2009.

[22] Tsai CL, Camargo CAJ. Methodological considerations, such as directed

acyclic graphs, for studying „acute on chronic" disease epidemiology: chronic obstructive pulmonary disease example. J Clin Epidemiol. 2009;62(9):982–90.

[23] Camelo LV, Giatti L, Chor D, Griep RH, Bensenor IM, Santos IS, i sur. Associations of life course socioeconomic position and job stress with carotid intima-media thickness. The Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). Soc Sci Med. 2015;141:91–9.

[24] Yang X, Li S, Pan L, Wang Q, Li H, Han M, i sur. Assessment of successful smoking cessation by psychological factors using the Bayesian network approach. Psychol Health Med. 2016;21(5):652–61.

[25] Moller SV, Hannerz H, Hansen AM, Burr H, Holtermann A. Multi-wave cohort study of sedentary work and risk of ischemic heart disease. Scand J Work Environ Health. 2016;42(1):43–51.

[26] Wang L, Bautista LE. Serum bilirubin and the risk of hypertension. Int J Epidemiol. 2015;44(1):142–52.

[27] Lucas P. Expert knowledge and its role in learning Bayesian Networks in medicine: an appraisal. U: Quaglini S, Barahona P, Andreassen S, ur. Artificial Intelligence in Medicine: 8th Conference on Artificial Intelligence in Medicine in Europe, AIME 2001 Cascais, Portugal, July 1–4, 2001, Proceedings. Berlin: Springer Berlin Heidelberg; 2001. Str. 156–66.

[28] Elliott JH, Grimshaw J, Altman R, Bero L, Goodman SN, Henry D, i sur. Informatics: Make sense of health data. Nature. 2015;527(7576):31–2.

[29] PLoS Medicine Editors. Can data sharing become the path of least resistance? PLoS Med. 2016;13(1):e1001949.

[30] Taichman DB, Backus J, Baethge C, Bauchner H, de Leeuw PW, Drazen JM, i sur. Sharing clinical trial data: a proposal from the International Committee of Medical Journal Editors. PLoS Med. 2016;13(1):e1001950.

[31] Davies J, Yudkin JS, Atun R. Liberating data: the crucial weapon in the fight against NCDs. Lancet Diabetes Endocrinol. 2016;4(3):197–8.

[32] Jackson CH, Best NG, Richardson S. Bayesian graphical models for regression on multiple data sets with different variables. Biostatistics. 2009;10(2):335–51.

[33] Zhao D, Weng C. Combining PubMed knowledge and EHR data to develop a weighted Bayesian network for pancreatic cancer prediction. J Biomed Inform. 2011;44(5):859–68.

[34] Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, i sur. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353:i3140.

[35] Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.

[36] Resche-Rigon M, White IR, Bartlett J, Peters SAE, Thompson SG, PROG-IMT Study Group. Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. Stat Med. 2013;32(28):4890–905.

[37] Jolani S, Debray TPA, Koffijberg H, van Buuren S, Moons KGM. Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. Stat Med. 2015;34(11):1841–63.

[38] Resche-Rigon M, White IR. Multiple imputation by chained equations for systematically and sporadically missing multilevel data. Stat Methods Med Res; 2016 Sep 19. doi: pii: 0962280216666564. [Epub ahead of print].

[39] The Fibrinogen Studies Collaboration. Systematically missing confounders in individual participant data meta-analysis of observational cohort studies. Stat Med. 2009;28(8):1218–37.

[40] Triantafillou S, Tsamardinos I. Constraint-based causal discovery from multiple interventions over overlapping variable sets. J Mach Learn Res. 2015;16:2147–205.

[41] Sabolić Pipinić I, Varnai VM, Turk R, Breljak D, Kezić S, Macan J. Low frequency of filaggrin null mutations in Croatia and their relation with allergic diseases. Int J Immunogenet. 2013;40:192–8.

[42] Gomzi M, Bobić J, Radošević-Vidaček B, Macan J, Varnai VM, Milković-Kraus S, i sur. Sick building syndrome: psychological, somatic, and environmental determinants. Arch Environ Occup Health. 2007;62(3):147–55.

[43] Macan J, Varnai VM, Maloča I, Kanceljak-Macan B. Increasing trend in atopy markers prevalence in a Croatian adult population between 1985 and 1999. Clin Exp Allergy. 2007;37(12):1756–63.

[44] World Health Organization. World Health Survey [Internet]. 2003 [pristupljeno 12.1.2017.]. Dostupno na: `http://www.who.int/healthinfo/survey/en/index.html`.

[45] Üstün T, Chatterji S, Mechbal A, Murray C, WHS Collaborating Groups. The World Health Surveys. U: Murray C, Evans D, ur. Health systems performance assessment: debates, methods, and empiricism. Geneva: World Health Organization; 2003. Str. 797–808.

[46] World Health Organization. WHO Multi-Country Studies Data Archive Central Data Catalog [Internet]. 2016 [pristupljeno 12.1.2017.]. Dostupno na: http://apps.who.int/healthinfo/systems/surveydata/index.php/catalog/central/about.

[47] OECD. Compendium of OECD well-being indicators [Internet]. OECD [pristupljeno 12.1.2017.]. Dostupno na: http://www.oecd.org/std/47917288.pdf.

[48] World Health Organization. World Health Survey: Guide to administration and question by question specifications [Internet]. World Health Organization [pristupljeno 12.1.2017.]. Dostupno na: http://www.who.int/healthinfo/survey/whslongversionguide.pdf?ua=1.

[49] World Health Organization. International guide for monitoring alcohol consumption and related harm. Geneva: World Health Organization; 2002.

[50] Hosseinpoor AR, Bergen N, Kunst A, Harper S, Guthold R, Rekve D, i sur. Socioeconomic inequalities in risk factors for non communicable diseases in low-income and middle-income countries: results from the World Health Survey. BMC Public Health. 2012;12(1):1–13.

[51] World Health Organization. Diet, nutrition and the prevention of chronic diseases. Geneva: World Health Organization; 1990.

[52] World Health Organization/Food and Agriculture Organization. Diet, nutrition and the prevention of chronic diseases. Geneva: World Health Organization; 2003.

[53] World Health Organization. Global recommendations on physical activity for health. Geneva: World Health Organization; 2010.

[54] WHOQOL Group. Development of the World Health Organization WHOQOL-BREF Quality of Life Assessment. Psychol Med. 1998;28(3):551–8.

[55] Lojk L. Eysenckov upitnik ličnosti - EPQ, priručnik. Ljubljana: Zavod SR Slovenije za produktivnost dela; 1984.

[56] Bobić J, Cvijetić S, Macan J. Personality and self-perception of physical and emotional health among first-year university students. Društvena istraživanja. 2015;24(2):219–37.

[57] Weider A, Wolff HG, Brodman K, Mittelmann B, Wechsler D. Priručnik za Cornell Index (hrvatska verzija). Jastrebarsko: Naklada Slap; 1997.

[58] Zung WK. A self-rating depression scale. Arch Gen Psychiatry. 1965;12(1):63–70.

[59] Cifuentes M, Sembajwe G, Tak S, Gore R, Kriebel D, Punnett L. The association of major depressive episodes with income inequality and the human development index. Soc Sci Med. 2008;67(4):529 –39.

[60] Loerbroks A, Herr RM, Subramanian S, Bosch JA. The association of asthma and wheezing with major depressive episodes: an analysis of 245 727 women and men from 57 countries. Int J Epidemiol. 2012;41(5):1436–44.

[61] Global Initiative for Asthma (GINA). The Global Strategy for Asthma Management and Prevention [Internet]. 2011 [pristupljeno 8.3.2017.]. Dostupno na: `https://tinyurl.com/zp5u2zd`.

[62] Dreborg S, Frew A. Position Paper: Allergen standardization and skin tests. Allergy. 1993;48:49–54.

[63] Aberer W, Kränke B. Measurement of IgE antibodies using liquid allergens–an inter-method and inter-laboratory quality assessment. Wien Klin Wochenschr. 2002;114(21–2):929–37.

[64] Brozek JL, Bousquet J, Baena-Cagnani CE, Bonini S, Canonica GW, Casale TB, i sur. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 Revision. J Allergy Clin Immunol. 2010;126(3):466 –76.

[65] Varnai VM, Ljubičić A, Prester L, Macan J. Exhaled breath condensate pH in adult Croatian population without respiratory disorders: how healthy a population should be to provide normative data? Arh Hig Rada Toksikol. 2009;60(1):87–97.

[66] Ljubičić Ćalušić A, Varnai VM, Čavlović AO, Šegvić Klarić M, Beljo R, Prester L, i sur. Respiratory health and breath condensate acidity in sawmill workers. Int Arch Occup Environ Health. 2013;86(7):815–25.

[67] Kovačić J, Varnai VM. A graphical model approach to systematically missing data in meta-analysis of observational studies. Stat Med. 2016;35(24):4443–58.

[68] Siegrist J, Wahrendorf M, von dem Knesebeck O, Jürges H, Börsch-Supan A. Quality of work, well-being, and intended early retirement of older employees—baseline results from the SHARE Study. Eur J Public Health. 2007;17(1):62–8.

[69] Lewis FI, Kratzer G, Pittavino M, Furrer R. R package abn: Modelling Multivariate Data with Additive Bayesian Networks [Internet]. 2016 [pristupljeno 18.12.2016]. Dostupno na: `https://CRAN.R-project.org/package=abn`.

[70] Lunn D, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput. 2000;10(4):325–37.

[71] Geweke J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. U: Bernardo JM, Berger J, Dawid AP, Smith JFM, ur. Bayesian Statistics 4. Oxford: Oxford University Press; 1992. Str. 169–93.

[72] R Core Team. R: A Language and Environment for Statistical Computing [Internet]. R Foundation for Statistical Computing [pristupljeno 13.1.2017.]. Dostupno na: `http://www.R-project.org`.

[73] Seltman HJ. R Package rube (Really Useful WinBUGS (or JAGS) Enhancer [Internet]. [pristupljeno 13.1.2017.].). Dostupno na: `http://www.stat.cmu.edu/~hseltman/rube/rubeManual.pdf`.

[74] Plummer M, Best N, Cowles K, Vines K. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News. 2006;6(1):7–11.

[75] Ide JS, Cozman FG. Random generation of Bayesian networks. U: Bittencourt G, Ramalho GL, ur. Advances in Artificial Intelligence. Berlin: Springer-Verlag; 2002. Str. 366–76.

[76] Scutari M. Learning Bayesian Networks with the bnlearn R Package. J Stat Softw. 2010;35(1):1–22.

[77] Triantafillou S. COmbINE [Internet]. 2016 [pristupljeno 13.1.2017.]. Dostupno na: `https://github.com/striantafillou/COmbINE`.

[78] Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80:27–38.

[79] Cazzoletti L, Ferrari M, Olivieri M, Verlato G, Antonicelli L, Bono R, i sur. The gender, age and risk factor distribution differs in self-reported allergic and non-allergic rhinitis: a cross-sectional population-based study. Allergy Asthma Clin Immunol. 2015;11(1):36.

[80] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, i sur. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.

[81] Kovačić J. Developing prediction models using a small number of datasets with overlapping variables. U: 21st Young Statisticians Meeting: Programme - Abstracts - Participants; 2016. .

[82] Grad R, Morgan WJ. Long-term outcomes of early-onset wheeze and asthma. J Allergy Clin Immunol. 2012;130(2):299–307.

[83] Sly RM. Changing prevalence of allergic rhinitis and asthma. Ann Allergy Asthma Immunol. 1999;82(3):233–48.

[84] Skoner DP. Allergic rhinitis: definition, epidemiology, pathophysiology, detection, and diagnosis. J Allergy Clin Immunol. 2001;108(1):S2–S8.

[85] Burgess JA, Walters EH, Byrnes GB, Matheson MC, Jenkins MA, Wharton CL, i sur. Childhood allergic rhinitis predicts asthma incidence and persistence to middle age: a longitudinal study. J Allergy Clin Immunol. 2007;120(4):863–9.

[86] Burgess JA, Dharmage SC, Byrnes MC G B ad Matheson, Gurrin LC, Wharton CL, Johns DP, i sur. Childhood eczema and asthma incidence and persistence: a cohort study from childhood to middle age. J Allergy Clin Immunol. 2008;122(2):280–5.

[87] Toelle BG, Xuan W, Peat JK, Marks GB. Childhood factors that predict asthma in young adulthood. Eur Respir J. 2003;23(1):66–70.

[88] Nemes S, Jonasson JM, Genell A, Steineck G. Bias in odds ratios by logistic regression modelling and sample size. BMC Med Res Methodol. 2009;9:56.

[89] Mortz CG, Andersen KE, Bindslev-Jensen C. Recall bias in childhood atopic diseases among adults in the Odense Adolescence Cohort Study. Acta Derm Venereol. 2015;95(8):968–72.

[90] Halken S, Host A. The lessons of noninterventional and interventional prospective studies on the development of atopic disease during childhood. Allergy. 2000;55(9):793–802.

[91] Burke W, Fesinmeyer M, Reed K, Hampson L, Carlsten C. Family history as a predictor of asthma risk. Am J Prev Med. 2003;24(2):160–9.

[92] Lundbäck B. Epidemiology of rhinitis and asthma. Clin Exp Allergy. 1998;28(Suppl 2):3–10.

[93] Raghavan D, Jain R. Increasing awareness of sex differences in airway diseases. Respirology. 2016;21(3):449–59.

[94] Zein JG, Erzurum SC. Asthma is Different in Women. Curr Allergy Asthma Rep. 2015;15(6):28.

[95] Bonds RS, Midoro-Horiuti T. Estrogen effects in allergy and asthma. Curr Opin Allergy Clin Immunol. 2013;13(1):92–9.

[96] Almqvist C, Worm M, Leynaert B. Impact of gender on asthma in childhood and adolescence: a GA2LEN review. Allergy. 2008;63(1):47–57.

[97] Ma J, Xiao L. Association of General and Central Obesity and Atopic and Nonatopic Asthma in US Adults. J Asthma. 2013;50(4):395–402.

[98] Leynaert B, Sunyer J, Garcia-Esteban R, Svanes C, Jarvis D, Cerveri I, i sur. Gender differences in prevalence, diagnosis and incidence of allergic and non-allergic asthma: a population-based cohort. Thorax. 2012;67(7):625–31.

[99] Kurukulaaratchy RJ, Karmaus W, Arshad SH. Gender and atopy influences on the natural history of rhinitis. Curr Opin Allergy Clin Immunol. 2012;12(1):7–12.

[100] Kurukulaaratchy RJ, Karmaus W, Raza A, Matthews S, Roberts G, Arshad SH. The influence of gender and atopy on the natural history of rhinitis in the first 18 years of life. Clin Exp Allergy. 2011;41(6):851–9.

[101] Keller T, Almqvist C, Kull I, Eller E, Christiansen ES, Wijga A, i sur. Sex-switch of allergic rhinitis prevalence during adolescence - pooled analyses of longitudinal European birth cohort data from MeDALL. Allergy. 2016;71(Suppl 102):27.

[102] Canuel M, Lebel G. Epidemiology of allergic rhinitis in Quebec: from a 2008 population-based survey. Chronic Dis Inj Can. 2014;34(2–3):163–8.

[103] Rhee CS, Wee JH, Ahn JC, Lee WH, Tan KL, Ahn S, i sur. Prevalence, risk factors and comorbidities of allergic rhinitis in South Korea: The Fifth Korea National Health and Nutrition Examination Survey. Am J Rhinol Allergy. 2014;28(2):e107–e114.

[104] Skaaby T, Husemoen LL, Thuesen BH, Jorgensen T, Linneberg A. Lifestyle-related factors and atopy in seven Danish population-based studies from different time periods. PLoS One. 2015;10(9):e0137406.

[105] Linneberg A, Nielsen NH, Madsen F, Frolund L, Dirksen A, Jorgensen T. Factors related to allergic sensitization to aeroallergens in a cross-sectional study in adults: The Copenhagen Allergy Study. Clin Exp Allergy. 2001;32(9):1409–17.

[106] Ferraz E, Garcia CAA, Bettiol HAs, Caldeira RDA, Cardoso VC, Arruda LK, i sur. Atopy risk factors at birth and in adulthood. J Pediatr (Rio J). 2011;87:336–42.

[107] Townsend EA, Miller VM, Prakash YS. Sex Differences and Sex Steroids in Lung Health and Disease. Endocr Rev. 2012;33(1):1–47.

[108] Baumann S, Lorentz A. Obesity - a promoter of allergy? Int Arch Allergy Immunol. 2013;162(3):205–13.

[109] Guh DP, Zhang W, Bansback N, Amarsi Z, Birmingham CL, Anis AH. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. BMC Public Health. 2009;9:88.

[110] Beuther DA, Sutherland ER. Overweight, obesity, and incident asthma: a meta-analysis of prospective epidemiologic studies. Am J Respir Crit Care Med. 2007;175(7):661–6.

[111] Ford ES. The epidemiology of obesity and asthma. J Allergy Clin Immunol. 2005;115(5):897–909.

[112] Shore SA. Obesity and asthma: possible mechanisms. J Allergy Clin Immunol. 2008;121(5):1087–93.

[113] Boulet LP. Obesity and atopy. Clin Exp Allergy. 2015;45(1):75–86.

[114] Chinn S. Obesity and asthma: evidence for and against a causal relation. J Asthma. 2003;40(1):1–16.

[115] Hersoug LG, Linneberg A. The link between the epidemics of obesity and allergic diseases: does obesity induce decreased immune tolerance? Allergy. 2007;62(10):1205–13.

[116] Schachter L, Salome C, Peat J, Woolcock A. Obesity is a risk for asthma and wheeze but not airway hyperresponsiveness. Thorax. 2001;56(1):4–8.

[117] Fenger RV, Gonzalez-Quintela A, Vidal C, Gude F, Husemoen LL, Aadahl M, i sur. Exploring the obesity-asthma link: do all types of adiposity increase the risk of asthma? Clin Exp Allergy. 2012;42(8):1237–45.

[118] Jarvis D, Chinn S, Potts J, Burney P. Association of body mass index with respiratory symptoms and atopy: results from the European Community Respiratory Health Survey. Clin Exp Allergy. 2002;32(6):831–7.

[119] Xu B, Jarvelin MR, Pekkanen J. Body build and atopy. J Allergy Clin Immunol. 2000;105(2 Pt 1):393–4.

[120] Vally H, Thompson PJ. Alcoholic drink consumption: a role in the development of allergic disease? Clin Exp Allergy. 2003;33(2):156–8.

[121] Vally H, Thompson PJ. Allergic and asthmatic reactions to alcoholic drinks. Addict Biol. 2003;8(1):3–11.

[122] Gonzalez-Quintela A, Vidal C, Gude F. Alcohol, IgE and allergy. Addict Biol. 2004;9(3–4):195–204.

[123] Scadding G. Alcohol consumption and allergic rhinitis. Clin Exp Allergy. 2008;38(7):1071–3.

[124] Linneberg A, Hertzum I, Husemoen LL, Johansen N, Jorgensen T. Association between alcohol consumption and aeroallergen sensitization in Danish adults. Clin Exp Allergy. 2006;36(6):714–21.

[125] Vidal C, Armisen M, Dominguez-Santalla MJ, Gude F, Lojo S, Gonzalez-Quintela A. Influence of alcohol consumption on serum immunoglobulin E levels in atopic and nonatopic adults. Alcohol Clin Exp Res. 2002;26(1):59–64.

[126] Assing K, Bodtger U, Linneberg A, Malling HJ, Poulsen LK. Association between alcohol consumption and skin prick test reactivity to aeroallergens. Ann Allergy Asthma Immunol. 2007;98(1):70–4.

[127] Bendtsen P, Gronbaek M, Kjaer SK, Munk C, Linneberg A, Tolstrup JS. Alcohol consumption and the risk of self-reported perennial and seasonal allergic rhinitis in young adult women in a population-based cohort study. Clin Exp Allergy. 2008;38(7):1179–85.

[128] Yoo Y, Perzanowski MS. Allergic sensitization and the environment: latest update. Curr Allergy Asthma Rep. 2014;14(10):465.

[129] Saulyte J, Regueira C, Montes-Martinez A, Khudyakov P, Takkouche B. Active or passive exposure to tobacco smoking and allergic rhinitis, allergic dermatitis, and food allergy in adults and children: a systematic review and meta-analysis. PLoS Med. 2014;11(3):e1001611.

[130] Feleszko W, Ruszczynski M, Jaworska J, Strzelak A, Zalewski BM, Kulus M. Environmental tobacco smoke exposure and risk of allergic sensitisation in children: a systematic review and meta-analysis. Arch Dis Child. 2014;99(11):985–92.

[131] Burke H, Leonardi-Bee J, Hashim A, Pine-Abata H, Chen Y, Cook DG, i sur. Prenatal and passive smoke exposure and incidence of asthma and wheeze: systematic review and meta-analysis. Pediatrics. 2012;129(4):735–44.

[132] Polosa R, Thomson NC. Smoking and asthma: dangerous liaisons. Eur Respir J. 2013;41(3):716–26.

[133] Toskala E, Kennedy DW. Asthma risk factors. Int Forum Allergy Rhinol. 2015;5(Suppl 1):S11–16.

[134] Eijkemans M, Mommers M, Draaisma JM, Thijs C, Prins MH. Physical activity and asthma: a systematic review and meta-analysis. PLoS One. 2012;7(12):e50775.

[135] Ford ES. Does exercise reduce inflammation? Physical activity and C-reactive protein among U.S. adults. Epidemiology. 2002;13(5):561–8.

[136] Vernon MK, Wiklund I, Bell JA, Dale P, Chapman KR. What do we know about asthma triggers? A review of the literature. J Asthma. 2012;49(10):991–8.

[137] Carson KV, Chandratilleke MG, Picot J, Brinn MP, Esterman AJ, Smith BJ. Physical training for asthma. Cochrane Database Syst Rev. 2013;30(9):CD001116.

[138] Gerald JK, Moreno FA. Asthma and depression: it's complicated. J Allergy Clin Immunol Pract. 2016;4(1):74–5.

[139] Gao YH, Zhao HS, Zhang FR, Gao Y, Shen P, Chen RC, i sur. The relationship between depression and asthma: a meta-analysis of prospective studies. PLoS One. 2015;10(7):e0132424.

[140] Jiang M, Qin P, Yang X. Comorbidity between depression and asthma via immune-inflammatory pathways: a meta-analysis. J Affect Disord. 2014;166:22–9.

[141] Chida Y, Hamer M, Steptoe A. A bidirectional relationship between psychosocial factors and atopic disorders: a systematic review and meta-analysis. Psychosom Med. 2008;70(1):102–16.

[142] Chen MH, Su TP, Chen YS, Hsu JW, Huang KL, Chang WH, i sur. Allergic rhinitis in adolescence increases the risk of depression in later life: a nationwide population-based prospective cohort study. J Affect Disord. 2013;145(1):49–53.

[143] Sanna L, Stuart AL, Pasco JA, Jacka FN, Berk M, Maes M, i sur. Atopic disorders and depression: findings from a large, population-based study. J Affect Disord. 2014;155:261–5.

[144] Chen CM, Tischer C, Schnappinger M, Heinrich J. The role of cats and dogs in asthma and allergy – a systematic review. Int J Hyg Environ Health. 2010;213(1):1 – 31.

[145] Konradsen JR, Fujisawa T, van Hage M, Hedlin G, Hilger C, Kleine-Tebbe J, i sur. Allergy to furry animals: New insights, diagnostic approaches, and challenges. J Allergy Clin Immunol. 2015;135(3):616 – 25.

[146] Singh M, Jaiswal N. Dehumidifiers for chronic asthma. Cochrane Database Syst Rev. 2013;13(6):CD003563.

[147] Zock JP, Jarvis D, Luczynska C, Sunyer J, Burney P. Housing characteristics, reported mold exposure, and asthma in the European Community Respiratory Health Survey. J Allergy Clin Immunol. 2002;110(2):285 – 92.

[148] Wright LS, Phipatanakul W. Environmental remediation in the treatment of allergy and asthma: latest updates. Curr Allergy Asthma Rep. 2014;14(3):419.

[149] Biagtan M, Viswanathan R, Bush RK. Immunotherapy for house dust mite sensitivity: where are the knowledge gaps? Curr Allergy Asthma Rep. 2014;14(12):482.

[150] Simpson A, Woodcock A, Custovic A. Housing characteristics and mite allergen levels: to humidity and beyond. Clin Exp Allergy. 2001;31(6):803–5.

[151] Hirsch T, Hering M, Bürkner K, Hirsch D, Leupold W, Kerkmann ML, i sur. House-dust-mite allergen concentrations (Der f 1) and mold spores in apartment bedrooms before and after installation of insulated windows and central heating systems. Allergy. 2000;55(1):79–83.

[152] Fairs A, Wardlaw AJ, Thompson JR, Pashley CH. Guidelines on ambient intramural airborne fungal spores. J Investig Allergol Clin Immunol. 2010;20(6):490–8.

[153] Köse S, Mandiracioğlu A, Çavdar G, Ulu Y, Senger SS. Prevalence of allergic diseases in Izmir Province, Turkey. Int Forum Allergy Rhinol. 2014;4(3):232–8.

[154] Global Initiative for Asthma (GINA). Global Strategy for Asthma Management and Prevention (updated 2016): Online Appendix [Internet].Global Initiative for Asthma (GINA) [pristupljeno 1.3.2017.]. Dostupno na: `http://ginasthma.org/wp-content/uploads/dlm_uploads/2016/02/WMSGINA-Appendix-2016-final.pdf`.

[155] Uphoff E, Cabieses B, Pinart M, Valdés M, Antó JM, Wright J. A systematic review of socioeconomic position in relation to asthma and allergic diseases. Eur Respir J. 2015;46(2):364–74.

[156] Lahey BB. Public health significance of neuroticism. Am Psychol. 2009;64(4):241–56.

[157] Costa PP, McCrae RR. Four ways five factors are basic. Pers Individ Dif. 1992;13(6):653–65.

[158] Goldberg LR. The structure of phenotypic personality traits. Am Psychol. 1993;48(1):26–34.

[159] Barnhofer T, Duggan DS, Griffith JW. Dispositional mindfulness moderates the relation between neuroticism and depressive symptoms. Pers Individ Dif. 2011;51(8):958–62.

[160] Graif Y, Goldberg A, Tamir R, Vigiser D, Melamed S. Skin test results and self-reported symptom severity in allergic rhinitis: the role of psychological factors. Clin Exp Allergy. 2006;36(12):1532–7.

[161] Goodwin RD, Castro M, Kovacs M. Major depression and allergy: does neuroticism explain the relationship? Psychosom Med. 2006;68(1):94–8.

[162] Loerbroks A, Apfelbacher CJ, Thayer JF, Debling D, Stürmer T. Neuroticism, extraversion, stressful life events and asthma: a cohort study of middle-aged adults. Allergy. 2009;64(10):1444–50.

[163] Cheng H, Deighton J, Wolpert M, Chapman BP, Kornilaki EN, Treglown L, i sur. Hay fever in childhood, traits Neuroticism and Conscientiousness as independent predictors of the occurrence of hay fever in adulthood. J Health Psychol. 2016;21(10):2367–75.

[164] Huovinen E, Kaprio J, Koskenvuo M. Asthma in relation to personality traits, life satisfaction, and stress: a prospective study among 11000 adults. Allergy. 2001;56(10):971–7.

[165] Loerbroks A, Li J, Bosch JA, Herr RM, Angerer P. Personality and risk of

adult asthma in a prospective cohort study. J Psychosom Res. 2015;79(1):13–7.

[166] Radošević-Vidaček B, Košćec A, Bakotić M, Macan J, Bobić J. Is atopy related to neuroticism, stress, and subjective quality of life? Arh Hig Rada Toksikol. 2009;60(1):99–107.

[167] Lu Y, Ho R, Lim TK, Kuan WS, Goh DY, Mahadevan M, i sur. Psychiatric comorbidities in Asian adolescent asthma patients and the contributions of neuroticism and perceived stress. J Adolesc Health. 2014;55(2):267–75.

[168] Axelsson M, Brink E, Lötvall J. A personality and gender perspective on adherence and health-related quality of life in people with asthma and/or allergic rhinitis. J Am Assoc Nurse Pract. 2014;26(1):32–9.

[169] Koca Kalkan I, Baccioglu Kavut A, Kalpaklioglu AF. Allergic rhinitis: can we identify nonadherence to therapy and its predictors easily in daily practice? J Investig Allergol Clin Immunol. 2013;23(5):315–22.

[170] Nurmatov U, Devereux G, Sheikh A. Nutrients and foods for the primary prevention of asthma and allergy: systematic review and meta-analysis. J Allergy Clin Immunol. 2011;127(3):724–733.e30.

[171] Garcia-Larsen V, Del Giacco SR, Moreira A, Bonini M, Charles D, Reeves T, i sur. Asthma and dietary intake: an overview of systematic reviews. Allergy. 2016;71(4):433–42.

[172] Seyedrezazadeh E, Moghaddam MP, Ansarin K, Vafa MR, Sharma S, Kolahdooz F. Fruit and vegetable intake and risk of wheezing and asthma: a systematic review and meta-analysis. Nutr Rev. 2014;72(7):411–28.

[173] Allan K, Kelly FJ, Devereux G. Antioxidants and allergic disease: a case of too little or too much? Clin Exp Allergy. 2010;40(3):370–80.

[174] Gudelj I, Mrkić Kobal I, Munivrana Škvorc H, Miše K, Vrbica Z, Plavec D, i sur. Intraregional differences in asthma prevalence and risk factors for asthma among adolescents in Split-Dalmatia County, Croatia. Med Sci Monit. 2012;18(4):PH43–50.

[175] Warm K, Hedman L, Lindberg A, L J. Allergic sensitization is age-dependently associated with rhinitis, but less so with asthma. J Allergy Clin Immunol. 2015;136(6):1559–65.e1–2.

[176] Tan DJ, Walters EH, Perret JL, Lodge CJ, Lowe AJ, Matheson MC, i sur. Age-of-asthma onset as a determinant of different asthma phenotypes in adults: a systematic review and meta-analysis of the literature. Expert Rev Respir Med. 2015;9(1):109–23.

[177] Langen U, Schmitz R, Steppuhn H. Häufigkeit allergischer Erkrankungen in Deutschland. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz. 2013;56(5):698–706.

[178] Gupta R, Sheikh A, Strachan DP, Anderson HR. Burden of allergic disease in the UK: secondary analyses of national databases. Clin Exp Allergy. 2004;34(4):520–6.

[179] Dorner T, Lawrence K, Rieder A, Kunze M. Epidemiology of allergies in Austria. Results of the first Austrian Allergy Report. Wien Med Wochenschr. 2007;157(11–12):235–42.

[180] Schröder PC, Li J, Wong GWK, Schaub B. The rural–urban enigma of allergy: what can we learn from studies around the world? Pediatr Allergy Immunol. 2015;26(2):95–102.

[181] D'Amato G, Vitale C, Lanza M, Molino A, D'Amato M. Climate change, air pollution, and allergic respiratory diseases: an update. Curr Opin Allergy Clin Immunol. 2016;16(5):434–40.

[182] von Mutius E, Vercelli D. Farm living: effects on childhood asthma and allergy. Nat Rev Immunol. 2010;10(12):861–8.

[183] Gasana J, Dillikar D, Mendy A, Forno E, Ramos Vieira E. Motor vehicle air pollution and asthma in children: a meta-analysis. Environ Res. 2012;117:36–45.

[184] Anderson HR, Favarato G, Atkinson RW. Long-term exposure to air pollution and the incidence of asthma: meta-analysis of cohort studies. Air Qual Atmos Health. 2013;6(1):47–56.

[185] Bowatte G, Lodge C, Lowe AJ, Erbas B, Perret J, Abramson MJ, i sur. The influence of childhood traffic-related air pollution exposure on asthma, allergy and sensitization: a systematic review and a meta-analysis of birth cohort studies. Allergy. 2015;70(3):245–56.

[186] Gowers AM, Cullinan P, Ayres JG, Anderson HR, Strachan DP, Holgate ST, i sur. Does outdoor air pollution induce new cases of asthma? Biological plausibility and evidence; a review. Respirology. 2012;17(6):887–98.

[187] Jang AS, Jun YJ, Park MK. Effects of air pollutants on upper airway disease. Curr Opin Allergy Clin Immunol. 2016;16(1):13–7.

[188] Gruzieva O, Gehring U, Aalberse R, Agius R, Beelen R, Behrendt H, i sur. Meta-analysis of air pollution exposure association with allergic sensitization in European birth cohorts. J Allergy Clin Immunol. 2014;133(3):767–76.e7.

[189] Didsbury MS, Kim S, Medway MM, Tong A, McTaggart SJ, Walker AM, i sur. Socio-economic status and quality of life in children with chronic disease: a systematic review. J Paediatr Child Health. 2016;52(12):1062–9.

[190] Silva N, Carona C, Crespo C, Canavarro MC. Quality of life in pediatric asthma patients and their parents: a meta-analysis on 20 years of research. Expert Rev Pharmacoecon Outcomes Res. 2015;15(3):499–519.

[191] Maspero J, Lee BW, Katelaris CH, Potter PC, Cingi C, Lopatin A, i sur. Quality of life and control of allergic rhinitis in patients from regions beyond western Europe and the United States. Clin Exp Allergy. 2012;42(12):1684–96.

[192] Tucak I, Nekić M. Neke odrednice zadovoljstva zdravljem odraslih osoba. Medica Jadertina. 2006;36(3–4):73–82.

[193] Kaliterna Lipovčan L, Prizmić-Larsen Z. Kvaliteta življenja, životno zadovoljstvo i osjećaj sreće u Hrvatskoj i europskim zemljama. U: Ott K, ur. Pridruživanje Hrvatske Europskoj uniji. Izazovi sudjelovanja. Zagreb: Institut za javne financije, Zaklada Friedrich Ebert; 2006. Str. 181–98.

[194] Clifton J, Gingrich N. Are citizens of the world satisfied with their health? Health Aff (Millwood). 2007;26(5):w545–w551.

[195] Kurpas D, Mroczek B, Sitko Z, Helicka D, Kuchar E. Quality of life and health care utilization in patients with chronic respiratory diseases. Adv Exp Med Biol. 2015;834:63–74.

11 Kratka biografija

Jelena Kovačić rođena je 26. ožujka 1987. godine u Splitu. Osnovnu i srednju školu završila je u Omišu. Studirala je na Matematičkom odsjeku Prirodoslovno-matematičkog fakulteta u Zagrebu, gdje je završila Sveučilišni preddiplomski studij Matematike (2005.–2008.) i Sveučilišni diplomski studij Matematičke statistike (2008.–2011.). Od 2011. godine radi kao znanstvena novakinja u Jedinici za medicinu rada i okoliša Instituta za medicinska istraživanja i medicinu rada u Zagrebu. Poslijediplomski doktorski studij Biomedicine i zdravstva na Medicinskom fakultetu Sveučilišta u Zagrebu upisala je 2011. godine. Od 2014. godine član je Izvršnog uredničkog odbora (urednica za statistiku) časopisa Arhiv za higijenu rada i toksikologiju. Autorica je dva izvorna znanstvena rada i jednog računalnog programskog paketa te koautorica dva izvorna znanstvena rada i jednog preglednog rada.