Procjena i različitost rizika od pristranosti u kliničkim pokusima s kirurškim intervencijama u Cochraneovim sustavnim pregledima

Barčot, Ognjen

Doctoral thesis / Disertacija

2021

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

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:171:988405

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

SVEUČILIŠTE U SPLITU
MEDICINSKI FAKULTET
UNIVERSITAS STUDIORUM SPALATENSIS
FACULTAS MEDICA

Repository / Repozitorij:

MEFST Repository

UNIVERSITY OF SPLIT

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

SVEUČILIŠTE U SPLITU
MEDICINSKI FAKULTET

Ognjen Barčot, dr. med.

Procjena i različitost rizika od pristranosti u kliničkim pokusima s kirurškim intervencijama u Cochraneovim sustavnim pregledima

Doktorska disertacija

Akademska godina: 2020./2021.

Mentorica:
prof. dr. sc. Livia Puljak, dr. med.

U Splitu, veljača 2021. godine

SVEUČILIŠTE U SPLITU
MEDICINSKI FAKULTET

Ognjen Barčot, dr. med.

Procjena i različitost rizika od pristranosti u kliničkim pokusima s kirurškim intervencijama u Cochraneovim sustavnim pregledima

Doktorska disertacija

Akademska godina: 2020./2021.

Mentorica:
prof. dr. sc. Livia Puljak, dr. med.

U Splitu, veljača 2021. godine

Voditelj rada: Ognjen Barčot, dr. med.
specijalist opće kirurgije, uže specijalnosti digestivne (abdominalne) kirurgije

Klinički Bolnički Centar Split
Klinika za kirurgiju
Odjel za abdominalnu-koloproktološku kirurgiju

Sveučilište u Splitu
Medicinski fakultet u Splitu
Poslijediplomski studij TRIBE – Translacijska istraživanja u biomedicini

ZAHVALA

Poštovanoj mentorici, prof. dr. sc. Liviji Puljak, zahvaljujem na ukazanom povjerenju, pomoći i podršci. Hvala na svemu naučenom, hvala na upornosti, hvala što ste uspjeli izvesti neizvedivo. I nadam se da s ovim doktoratom naša suradnja nije završena, nego tek počinje.

Mojem kolegi, Matiji Boriću, zahvaljujem na svemu. Malo je listova i vremena za nabrojati svaku njegovu zaslugu. I zato se nadam da će naša suradnja, a prije svega prijateljstvo trajati još dugo.

Naposlijetku, hvala mojoj obitelji, mojoj Sanji na razumijevanju i potpori te mojoj djeci, Dorotei i Mislavu. Hvala vam za svo vrijeme koje nisam proveo s vama. Mislim da ga nisam potratio.

SADRŽAJ

POPIS OZNAKA I KRATICA 2
UVOD 3
1. Medicina utemeljena na dokazima 3
2. Hijerarhija dokaza u medicini 8
3. Sustavni pregled 12
4. Cochrane 14
5. Procjena rizika od pristranosti 16
CILJEVI I HIPOTEZE 24
Ciljevi 24
Hipoteze 24
METODE 25
1. Uzorak i zajedničko prikupljanje podataka 25
2. Metode rada o prikladnosti procjena Cochraneovih autora za domenu rizika pristranosti o generiranju nasumičnog redoslijeda 28
3. Metode rada o prikladnosti procjena Cochraneovih autora za domenu zasljepljivanja ispitanika i osoblja 33
4. Metode rada o prikladnosti procjena Cochraneovih autora za domenu rizika pristranosti o zasljepljivanju osoba koje bilježe ishode 40
5. Metode rada o istraživanju o kirurškim intervencijama 42
REZULTATI 45
1. Rezultati prvog rada – prikladnost procjena Cochraneovih autora za domenu rizika pristranosti o generiranju nasumičnog redoslijeda 45
2. Rezultati drugog rada – prikladnost procjena Cochraneovih autora za domenu zasljepljivanja ispitanika i osoblja 47
3. Rezultati trećeg rada – prikladnost procjena Cochraneovih autora za domenu rizika pristranosti o zasljepljivanju osoba koje bilježe ishode 55
4. Rezultati četvrtog rada - istraživanje o kirurškim intervencijama 61
RASPRAVA 69
1. Prikladnost procjena Cochraneovih autora za domenu rizika pristranosti o generiranju nasumičnog redoslijeda 69
2. Prikladnost procjena Cochraneovih autora za domenu zasljepljivanja ispitanika i osoblja 72
3. Prikladnost procjena Cochraneovih autora za domenu rizika pristranosti o zasljepljivanju osoba koje bilježe ishode 74
4. Prikladnost procjena Cochraneovih autora u istraživanju o kirurškim intervencijama 78
ZAKLJUČAK 82
KRATKI SAŽETAK 84
TITLE AND ABSTRACT OF DOCTORAL THESIS 86
DODATNI MATERIJALI 88
POPIS LITERATURE 96
BIOGRAFIJA 106

POPIS OZNAKA I KRATICA

Oznaka / kratica	**Značenje**
ASPS	*American Society of Plastic Surgeons*
CDSR	*Cochrane Database of Systematic Reviews*
CEBM	*Centre for Evidence Based Medicine*
CMAJ	*Canadian Medical Association Journal*
CONSORT	*Consolidated Standards of Reporting Trials*
CSR	*Cochrane Systematic Review*
DOI	*Digital Object Identifier*
EBM	*Evidence based medicine*
GRADE	*Grading of Recommendations, Assessment, Development and Evaluation*
IVRS	*Interactive Voice Response System*
JAMA	*Journal of the American Medical Association*
LOC	*Length of Comment*
MA	*Meta-analysis*
OSF	*Open-Source Framework*
PLS	*Plain Language Summary*
QPMC	*Quality and Performance Measurement Committee*
RCT	*Randomised Controlled Trial*
SNOSE	*Sequentially Numbered, Opaque, Sealed Envelopes*
SR	*Systematic Review*
STROBE	*STREngthening the Reporting of Observational Studies in Epidemiology*
URL	*Uniform Resource Locator*
VBA	*Visual Basic for Applications*

UVOD

1. Medicina utemeljena na dokazima

"Medicina utemeljena na dokazima (engl. evidence-based medicine – EBM) savjesna je, eksplicitna i razborita upotreba najboljih trenutnih dokaza u donošenju odluka o liječenju svakog pojedinog bolesnika. Praksa medicine utemeljene na dokazima jest integriranje vlastitog kliničkog iskustva s najboljim dostupnim vanjskim kliničkim dokazima iz sustavnih istraživanja."(1)

1.1. Definicija medicine utemeljene na dokazima

Navedeni citat dr. Davida Sacketta, u doslovnom prijevodu s engleskog jezika najbolje sažima definiciju medicine utemeljene na dokazima. Pod vlastitim kliničkim iskustvom podrazumijeva se stručnost i prosudba koju pojedinci stječu kliničkom praksom. Veliko iskustvo ogleda se na mnogo načina, ali posebno u svrsishodnom donošenju dijagnoza te u promišljenijoj identifikaciji komorbiditeta, kao i suosjećajnosti prema potrebama, pravima i preferencijama pojedinih pacijenata prilikom donošenja kliničkih odluka o njihovoj dijagnostici i liječenju (1).

Pod najboljim dostupnim vanjskim kliničkim dokazima podrazumijevaju se klinički relevantna istraživanja, često iz temeljnih medicinskih znanosti, a posebno klinička istraživanja o točnosti i preciznosti dijagnostičkih testova usmjerena na bolesnika. Ovoj grupi pripadaju i testovi osjetljivosti i specifičnosti prognostičkih biljega, testovi učinkovitosti i sigurnosti terapijskih, rehabilitacijskih i preventivnih postupaka. Vanjski klinički dokazi mogu poništiti prethodno prihvaćene dijagnostičke testove i načine liječenja te ih zamijeniti novim, snažnijim, preciznijim, učinkovitijim i sigurnijim (1).

Medicinska praksa trebala bi biti integracija najboljih istraživačkih dokaza s kliničkom stručnošću prilagođena pojedinom bolesniku. Medicina utemeljena na dokazima interdisciplinarni je pristup koji koristi tehnike iz znanosti, inženjerstva, biostatistike i epidemiologije. Ona je ideal po kojem bi zdravstveni radnici trebali savjesno i razumno koristiti trenutno najbolje dokaze u svojoj svakodnevnoj praksi (1).

Praktično gledano, najbolji dokazi dolaze iz sustavnih pregleda medicinske literature o određenim kliničkim temama (sinteza dokaza). Dokazi se zatim prevode u praksu od strane liječnika koji odabiru mogućnosti liječenja za određene slučajeve na temelju najboljih

istraživanja, sklonosti i individualnih karakteristika bolesnika. Praktičari medicine utemeljene na dokazima sudjeluju u cjeloživotnom učenju i predani su kontinuiranom obrazovanju od strane stručnjaka i zajednice pacijenata (2).

1.2. Povijest medicine utemeljene na dokazima

U proljeće 1990. na Sveučilištu McMaster, dr. Gordon Guyatt, internist, predstavio je novi koncept koji je nazvao "znanstvena medicina" (engl. *scientific medicine*). Htio je opisati novu metodu podučavanja medicine uz bolesnički krevet. Izgrađen je na temeljima koje je postavio njegov mentor dr. David Sackett. Međutim, njegove prijedloge kolege nisu spremno prihvatili. Implikacija da trenutne kliničke odluke nisu baš znanstveno potvrđene, iako vjerojatno istinita, nije im bila prihvatljiva. Guyatt se zatim pojavio s novim naslovom kojim je opisao temeljni okvir specijalističkog, ali i općemedicinskog usavršavanja: "Medicina utemeljena na dokazima" (engl. *EBM*). Taj pojam se po prvi puta u literaturi pojavio u uvodniku *ACP Journal Club* 1991. godine (3).

Iako je pojam uveden 1991. godine, temelj te nove strategije postignut je godinama rada mnogih drugih. Koncept i poticaj za EBM mogu se pripisati sve većoj svijesti o slabostima standardne kliničke prakse i njihovom utjecaju na kvalitetu, ali i cijenu skrbi o pacijentima. Klinička praksa u povijesti se smatrala "medicinskom umjetnošću" dok su stručno mišljenje, iskustvo i autoritarna prosudba bili temelj za donošenje odluka. Korištenje znanstvene metodologije, kao u biomedicinskim istraživanjima, i statističke analize, kao u epidemiologiji, bile su rijetke u svijetu medicine. Povijesno utvrđena hijerarhija i indoktrinirano nepovjerenje u ove discipline predstavljale su prepreku za implementaciju (4). Međutim, nekoliko događaja i znanstvenika u različitim dijelovima svijeta tijekom 1960-ih otvorilo je put EBM-u (5).

Nije se radilo samo o liječnicima i predavačima na medicinskim fakultetima sveučilišta Harvard, Stanford ili Johns Hopkins poput pionira Suzanne i Roberta Fletchera koji su 1982 objavili udžbenik koji opisuje znanstvenu osnovu za kliničku skrb, *Clinical Epidemiology: The Essentials*. U tom istom razdoblju Alvan Feinstein, matematičar-liječnik, predložio je da se nesigurnost u odlukama može svesti na najmanju razinu upotrebom novog načina razmišljanja – u polju temeljnih znanosti. Najraniji rad započeo je u bolnici za reumatsku groznicu u New Yorku, gdje je prepoznao da se razlikovanje benignih od patoloških srčanih šumova mora koristiti kao znanstveni kriterij za dijagnozu, suprotno isključivo kliničkom autoritetu (6). Njegova uspješna klasifikacija bolesti dovela je do poboljšanih ishoda i, ironično, zatvaranja bolnice za reumatsku groznicu zbog nedostatka pacijenata. Predložio je pojam "klinička

epidemiologija" (engl. *clinical epidemiology*) (7) i tako detaljno opisao novu disciplinu medicinske nastave. Feinstein je tako premostio svijet epidemiologije i medicinskih istraživanja, koji su dugo bili odvojeni, potičući tako korisnost medicinskih istraživanja šire od svojih tradicionalnih dosega (7).

Klinička epidemiologija postala je formalni studij prvo na novom medicinskom fakultetu Sveučilišta McMaster 1967. kada je uveden i novi Odjel za kliničku epidemiologiju i biostatistiku. Novi odjel dodijeljen je dr. Davidu Sackettu koji je dijelio viziju da je klinička epidemiologija primjena epidemioloških i biometrijskih metoda u proučavanju dijagnostičkog i terapijskog procesa, od strane liječnika koji pruža izravnu njegu pacijenta, a sve u svrhu poboljšanja zdravlja (8).

Sackett i njegovi kolege uvidjeli su potrebu ne samo podučavati metode razumijevanja literature, već i poučavati liječnike kako primijeniti nove informacije pri bolesničkom krevetu. Izdali su 1981. godine seriju članaka u *Canadian Medical Association Journal (CMAJ)* i tu definirali pojam "kritička procjena" (9).

Gordon Guyatt i njegovi kolege usavršavali su rad medicinskog podučavanja utemeljenog na dokazima na Sveučilištu McMaster tijekom 1990-ih. Zatim su surađivali s američkim kolegama i formirali međunarodnu radnu skupinu za EBM (10). Njihov rad u primjeni kritičke procjene doveo je do drastičnih promjena u kliničkoj praksi. S druge strane, sama procjena dovela je do zaključaka da su i sami članci u CMAJ-u na određen način ograničeni. Konkretno, ti su se članci prvenstveno fokusirali na kvalitetu dokaza - manje na primjenu dokaza na određeni klinički scenarij. Literatura je uvijek imala različite razine dokaza, ali nije samo stvar u procjeni kvalitete, problem nastaje kada se primjenjuju ponekad jedini postojeći, a ipak neoptimalni, dokazi. Štoviše, tim je člancima nedostajalo dosljednosti u objašnjenju pristranosti, valjanosti i preciznosti, što je čitatelja ostavilo potencijalno zbunjenog njihovim značenjem. Potreba za stvaranjem "korisničkog", a ne samo "čitateljskog" vodiča postala je očita tim pionirima. Stoga je u časopisu Američkog liječničkog udruženja (engl. *Journal of the American Medical Association - JAMA*) uspostavljen koncept Vodiča za korisnike. Vodič je nastao na temeljima dvadeset i pet članaka objavljenih od 1993. do 2000., s ciljem da se liječnicima praktičarima pomogne da razumiju i primijene literaturu na određenog pacijenta. Vodič za korisnike za medicinsku literaturu od 2015. godine slavi svoje treće izdanje (engl. *Users' Guides to the Medical Literature: A Manual of Evidence-Based Clinical Practice* (11)).

Danas je EBM postao temeljni princip podučavanja kliničke medicine, a u skladu s tim uvedene su promjene u medicinskoj edukaciji, kojima je cilj pomoći budućim liječnicima ovladati ključnim vještinama za primjenu EBM-a (12).

1.3. Kritike medicini utemeljenoj na dokazima

Iako je EBM sada formalno etabliran na sveučilištima i u kliničkoj praksi, postoje i značajne kritike prema tom pristupu medicini. Kritičari tvrde da EBM-u nedostaje korisnost na nekoliko razina. Neki tvrde da transformira složeni proces kliničkog odlučivanja - koji uključuje prikupljanje podataka, godine medicinskog znanja, iskustva i pametnu intuiciju - u algoritamsku vježbu koja nije individualizirana za određene kliničke scenarije i zbog toga podliježe pogreškama u skrbi o pacijentu. Čak je i sam Feinstein kritizirao radove o EBM-u sa Sveučilišta McMaster. Argument da EBM uključuje "najbolje dostupne vanjske kliničke dokaze iz sustavnog pretraživanja"(1) zahtijeva razumijevanje onoga što čini takav najbolji dokaz. Za pitanja o liječenju, randomizirani klinički pokusi (engl. *randomised controlled trial – RCT*) i sustavni pregled (engl. *systematic review – SR*) i meta-analiza (engl. *meta-analysis – MA*) su zlatni standard za EBM, u odnosu na ne-eksperimentalne pristupe. Kao primjere, Feinstein je naveo da su i inzulin za dijabetičku acidozu i penicilin za bakterijski endokarditis opisani kroz pojedinačne članke, a ne sustavnim pregledima literature. Tvrdio je da se zagovornici EBM-a previše oslanjaju na RCT-ove, za koje je naveo da su jednostavno usporedba jednog tretmana s drugim, a ne neki apsolutni oblik istine. Oslanjajući se na te epidemiološke alate, EBM ne uključuje "meke" podatke koje kliničari koriste za formuliranje dijagnoza i liječenja. Ti "mekani" podaci uključuju vrstu i težinu simptoma i brzinu rasta bolesti (13). Isto tako, socijalni i politički kontekst u kojem pacijenti žive nisu podjednako obrađeni u EBM-u. Na kraju, kritike EBM-a navode potencijal zlouporabe oznake „najbolji raspoloživi dokazi". Donositelji zdravstvenih politika, vladini i nevladini (privatni) investitori mogu preusmjeriti i opravdati financijski povrat na temelju „najboljih dostupnih dokaza" i marginalizirati praksu koja nije u skladu s tim standardima (1).

Još jedna kritika uključuje i poteškoće s kojima se kliničari suočavaju u praćenju ukupnog medicinskog napretka objavljenog u znanstvenim časopisima. To je očito iz usporedbe vremena potrebnog za čitanje (za opću medicinu, 1996. godine, bilo je potrebno proučiti 19 članaka dnevno, 365 dana u godini (14)) s dostupnim vremenom (medicinski savjetnici u Oxfordu odvajali su oko sat vremena tjedno (15), i to na temelju samoprocjene). Danas se glavne kritike i ograničenja EBM mogu sažeti i kategorizirati u pet tema: oslanjanje na empirizam (RCT), uska i nejasna definicija jakih dokaza, nedostatak dokaza o djelotvornosti, ograničena korisnost za pojedinačne bolesnike i prijetnja autonomiji odnosa liječnik – pacijent (16).

Međutim, strahovanje da će investitori i menadžeri zlouporabiti medicinu utemeljenu na dokazima kako bi smanjili troškove zdravstvene zaštite ukazuje na temeljno nerazumijevanje

njezinih financijskih posljedica. Liječnici koji se bave medicinom koja se temelji na dokazima pronaći će i primijeniti najučinkovitije intervencije kako bi povećali kvalitetu života za pojedine pacijente; to može povećati, a ne smanjiti troškove njihove skrbi. Isto tako EBM nipošto nije ograničena samo na RCT-ove i meta-analize upravo zbog stalnog traženja najboljih vanjskih dokaza pomoću kojih možemo odgovoriti na naša klinička pitanja. Za saznati točnost dijagnostičkog testa, metodologija EBM-a nalaže pronalaženje ili izradu odgovarajućeg presječnog istraživanja o pacijentima za koje se sumnja da imaju odgovarajući poremećaj, a ne samo RCT. Za pitanje o prognozi, potrebne su odgovarajuća istraživanja praćenja ishoda bolesnika uključenih u istraživanje u ranoj fazi kliničkog tijeka njihove bolesti. Kada se postavljaju pitanja o terapiji, trebalo bi pokušati izbjeći ne-eksperimentalne pristupe, jer oni rutinski dovode do lažno-pozitivnih zaključaka o učinkovitosti. A budući da je za RCT, a posebno za sustavni pregled (SR) RCT-a, mnogo vjerojatnije da će biti informativni, onda je utoliko manje vjerojatno da će nas zavesti. Stoga su SR postali zlatnim standardom za ocjenjivanje donosi li intervencija više koristi, nego štete. Štoviše, neka pitanja o terapiji, potpuno opravdano, ne zahtijevaju provedbu RCT-a (uspješne intervencije za inače smrtonosna stanja) ili ne mogu čekati provođenje istraživanja. A ako nije proveden RCT zbog teške situacije pojedinih bolesnika, onda se predlaže slijediti trag do sljedećih najboljih vanjskih dokaza i od tamo raditi (1).

Analiza EBM-a prema ovim temama dovodi do zaključka da EBM može biti koristan alat, ali ima ozbiljne nedostatke kada se izolirano koristi u praksi liječenja pojedinog bolesnika. Suvremena medicina mora težiti uravnoteženju izuzetno složenog niza prioriteta. Nesavršenost EBM-a, međutim, možda i nije njegova mana jer se koncept i dalje razvija kao odgovor na njegove kritike. Da bi bila učinkovita pomoć u postizanju ove ravnoteže, teorija i praksa EBM-a moraju se proširiti tako da uključuju nove metode ustroja i integracije istraživanja. Razvijene su strategije za razvoj visokokvalitetnih istraživanja i ako se tih smjernica u izradi istraživanja drži većina istraživača onda bi to trebalo dovesti do valjanijih i pouzdanijih rezultata. U budućnosti, uzimajući u obzir da će se i dalje pojavljivati novi načini dobivanja zdravstvenih podataka, svijet EBM-a i kliničkih istraživanja vjerojatno će se promijeniti. Krajnji cilj, međutim, mora ostati isti: poboljšanje zdravstvenih ishoda za bolesnike (17).

2. Hijerarhija dokaza u medicini

Temelj EBM-a je hijerarhijski sustav razvrstavanja dokaza. Ta je hijerarhija poznata kao razina dokaza. Zdravstveni radnici se potiču da pronađu najvišu razinu dokaza kako bi odgovorili na klinička pitanja. Prije nego što se to može postići, važno je razumjeti povijest iza tih razina dokaza i kako ih treba tumačiti (18).

U izvješću Kanadske radne skupine za periodične zdravstvene preglede (engl. *Canadian Task Force on the Periodic Health Examination*) 1979. godine po prvi puta izvorno su opisane razine dokaza (19). Svrha izvješća bila je razviti preporuke o periodičnim zdravstvenim pregledima i te preporuke temeljiti na dokazima u medicinskoj literaturi. Autori su razvili sustav ocjenjivanja dokaza (Tablica 1) pri određivanju učinkovitosti određene intervencije (19).

Tablica 1 - Razina dokaza prema Kanadskoj radnoj skupini za periodične zdravstvene preglede (19).

Razina	**Vrsta dokaza**
I	Najmanje 1 RCT uz pravilnu randomizaciju
II.1	Dobro osmišljeno kohortno istraživanje ili istraživanje parova
II.2	Usporedbe vremenskih serija ili dramatični rezultati nekontroliranih istraživanja
III	Mišljenja stručnjaka

Prilikom ocjenjivanja preporuka uzeti su u obzir raspoloživi dokazi. Na primjer, najviša preporuka za ocjenu dala se ako postoje dobri dokazi koji podupiru preporuku da se neko stanje uključi u periodični zdravstveni pregled. Razine dokaza nadalje je opisao i proširio Sackett u članku o razinama dokaza o antitrombotičkim agensima iz 1989. godine (Tablica 2) (20).

Oba sustava postavljaju RCT-ove na najvišu razinu, a serije slučajeva ili mišljenja stručnjaka na najnižu razinu. Hijerarhije rangiraju istraživanja prema vjerojatnosti pristranosti. RCT-ovi se smatraju najvišom razinom jer su ustrojeni tako da budu nepristrani i imaju manji rizik od sustavnih pogrešaka. Na primjer, nasumičnim razvrstavanjem ispitanika u dvije ili više skupina liječenja RCT-ovi također randomiziraju zbunjujuće čimbenike koji mogu rezultirati pristranosti. Serija slučaja ili mišljenje stručnjaka često su pristrani zbog autorovog iskustva ili mišljenja i nema kontrole zbunjujućih čimbenika (20).

Tablica 2 - Razina dokaza prema Sackettu 1989 (20).

Razina	**Vrsta dokaza**
I	Veliki RCT-ovi s jasnim rezultatima
II	Mali RCT-ovi s nejasnim rezultatima
III	Kohortna istraživanja i istraživanje parova
IV	Povijesna kohortna istraživanja ili prikaz slučajeva
V	Serije slučajeva, istraživanja bez kontrola

Od uvođenja razine dokaza, više organizacija i časopisa prihvatilo je različite varijacije tog sustava klasifikacije. Razne specijalnosti često postavljaju različita pitanja, a prepoznato je da vrstu i razinu dokaza treba prilagoditi u skladu s tim. Također, istraživačka pitanja podijeljena su u različite kategorije: liječenje, prognoza, dijagnoza i ekonomska analiza/odluka. Na primjer, Tablica 3 prikazuje razine dokaza koje je za prognozu razvilo Američko društvo plastičnih kirurga (engl. *American Society of Plastic Surgeons - ASPS*) (21), a

Tablica 4 razine koje je za liječenje razvio Centar za medicinu utemeljenu na dokazima (engl. *Centre for Evidence Based Medicine - CEBM*) (22).

Tablica 3 - Razina dokaza za prognostička istraživanja prema ASPS (21)

Razina	**Vrsta dokaza**
I	Kvalitetno prospektivno kohortno istraživanje s odgovarajućom snagom ili sustavnim pregledom tih istraživanja
II	Prospektivna kohorta manje kvalitete, retrospektivno kohortno istraživanje, neliječene kontrole iz RCT-a ili sustavni pregled tih istraživanja
III	Istraživanje parova ili sustavni pregled tih istraživanja
IV	Serija slučajeva
V	Mišljenje stručnjaka; prikaz slučaja ili klinički primjer; dokazi koji se temelje na fiziologiji, istraživanju na klupi ili “prvim principima”

Tablica 4 - Razina dokaza za istraživanja o terapiji prema CEBM (22)

Razina	**Vrsta dokaza**
1A	Sustavni pregled (s homogenim rezultatima) RCT-a
1B	Pojedinačni RCT (s uskim intervalima pouzdanosti)
1C	Istraživanja „sve ili ništa” (engl. *all or none*)
2A	Sustavni pregled (s homogenim rezultatima) kohortnih istraživanja
2B	Pojedinačna kohortna istraživanja (uključujući RCT niske kvalitete, npr. <80% praćenja)
2C	Istraživanje „ishoda”; ekološka istraživanja
3A	Sustavni pregled (s homogenim rezultatima) istraživanja parova
3B	Pojedinačna istraživanja parova
4	Serije slučajeva (usto i nekvalitetna kohorta; nekvalitetna istraživanja parova)
5	Mišljenje stručnjaka bez izričite kritičke ocjene ili utemeljeno na fiziološkim istraživanjima, temeljnim istraživanjima ili “prvim načelima”

Tablica 3 i Tablica 4 ističu vrste istraživanja koje su prikladne za pojedinu vrstu kliničkog pitanja (prognoza naspram terapije) i kako se kvaliteta podataka uzima u obzir pri određivanju razine. Primjerice, RCT nisu prikladni kada se gleda prognoza bolesti. Pitanje u tom slučaju je: “Što će se dogoditi ako uopće ne učinimo ništa?”. Budući da prognostičko pitanje ne uključuje usporedbu liječenja, najviše dokaza dobit će se iz kohortnog istraživanja ili sustavnog pregleda kohortnih istraživanja. Razine dokaza također uzimaju u obzir kvalitetu podataka. Na primjer, u razinama dokaza CEBM-a, loše ustrojeni RCT imaju istu razinu dokaza kao i kohortno istraživanje (20).

S vremenom se mijenjao i sustav ocjenjivanja koji daje snagu preporuka na temelju dokaza. Tablica 5 prikazuje preporuke za ocjenjivanje koje je razvio ASPS i koristio do 2016. Sustav ocjenjivanja pruža važnu komponentu u medicini koja se temelji na dokazima i pomaže u donošenju kliničkih odluka. Na primjer, daje se snažna preporuka kada postoje dostupni dokazi razine I i dosljedni dokazi iz istraživanja razine II, III i IV. Sustav ocjenjivanja ne umanjuje vrijednost dokaza niže razine prilikom odlučivanja o preporukama ako su rezultati dosljedni (23).

Tablica 5 – Preporuke ASPS za kliničku praksu do 2016 (23)

Ocjena	**Opis**	**Razina kvalifikacije dokaza**	**Utjecaj i preporuka za praksu**
A	Snažna preporuka	Dokazi razine I ili dosljedni rezultati iz više istraživanja razina II, III ili IV	Kliničari bi se trebali pridržavati snažne preporuke, osim ako ne postoji jasno i uvjerljivo obrazloženje za drugačiji pristup
B	Preporuka	Dokazi i rezultati razine II, III ili IV uglavnom su dosljedni	Općenito, kliničari bi se trebali pridržavati preporuka, ali trebali bi biti oprezni prema novim informacijama i propitati sklonosti pacijenta
C	Opcija	Dokazi razine II, III ili IV, ali rezultati su nedosljedni	Kliničari bi trebali biti fleksibilni u donošenju odluka u vezi s odgovarajućom praksom, iako mogu postaviti granice prema drugim pristupima; želje pacijenta trebale bi imati značajnu ulogu u odlučivanju
D	Opcija	Dokazi razine V: malo ili nimalo sustavnih empirijskih dokaza	Kliničari bi trebali razmotriti sve mogućnosti u svom donošenju odluka i pratiti nove objavljene dokaze koji pojašnjavaju omjer koristi i štete; želje pacijenta trebale bi imati značajnu ulogu u odlučivanju

Povjerenstvo za kvalitetu i izvedbu (engl. *Quality and Performance Measurement Committee* - *QPMC*) ASPS-a izglasalo je promjenu metodologije koja se koristi za razvoj smjernica kliničke prakse za ocjenjivanje preporuka te se od 2016. godine koristi sustav GRADE (engl. *Grading of Recommendations, Assessment, Development and Evaluation*) (24, 25).

GRADE sustav razvoja smjernica ima brojne prednosti nad dosadašnjim sustavima: jasno razdvaja kvalitetu dokaza i snagu preporuka; eksplicitno procjenjuje važnost ishoda alternativnih strategija; inzistira na sveobuhvatnosti kriterija za ocjenu kvalitete dokaza; ima transparentan postupak prelaska s dokaza na preporuke; pragmatično tumačenje jakih naspram slabih preporuka za kliničare; koristan je za sustavne preglede kao i smjernice (24).

Da bi se postigla transparentnost i jednostavnost, GRADE sustav klasificira kvalitetu dokaza na jednu od četiri razine - visoku, umjerenu, nisku i vrlo nisku. Neke organizacije koje koriste sustav GRADE odlučile su kombinirati niske i vrlo niske kategorije. Dokazi temeljeni na RCT započinju kao visokokvalitetni dokazi, ali kvaliteta (vjerodostojnost) dokaza može se smanjiti iz nekoliko razloga: ograničenja istraživanja; neusklađenost rezultata; neizravnost dokaza; nepreciznost; izvješćivanje o riziku od pristranosti (engl. *Risk of Bias – RoB*) (24).

3. Sustavni pregled

Sustavni pregledi (engl. *systematic review – SR*) koriste se za informiranje pri donošenju odluka u mnogim disciplinama, a usko ih se povezuje s medicinom utemeljenom na dokazima. SR pružaju iscrpan sažetak trenutne literature relevantne za određeno istraživačko pitanje. SR koristi sustavne i ponovljive metode za pronalaženje, odabir i kritičku procjenu svih relevantnih istraživanja te za prikupljanje i analizu podataka iz istraživanja koja su uključena u pregled (26).

Mnogi SR uključuju kvantitativnu analizu, meta-analizu (engl. *meta-analysis – MA*) dostupnih podataka pa ih se (SR i MA) učestalo i krivo poistovjećuje. Sustavni pregledi kvantitativnih podataka ili pregledi kombiniranih metoda koriste meta-analizu (kao statističku tehniku) za kombiniranje rezultata prihvatljivih istraživanja, a sve kako bi se povećala statistička značajnost rezultata. Međutim, postoje i SR koji ne uključuju MA jer MA nije bila moguća zbog nedostatka podataka ili nije primjerena zbog statističke ili kliničke heterogenosti uključenih istraživanja (27, 28).

SR koriste rigorozan i transparentan pristup za sintezu istraživanja, s ciljem procjene i, gdje je to moguće, smanjenja pristranosti u rezultatima. Iako postoji više vrsta SR, glavne faze pregleda mogu se sažeti u pet faza (29, 30):

i) Definiranje istraživačkog pitanja: pitanje na koje treba odgovoriti i slaganje s objektivnom metodom potrebni su za pravilnu izradu korisnog SR. Preporučuje se objavljivanje protokola pregleda prije početka istraživanja, kako bi se smanjio rizik od neplaniranog dupliciranja istraživanja i kako bi se osigurala dosljednost između metodologije i protokola (31).

ii) Traženje relevantnih izvora podataka: ustrojava se strategija pretraživanja za pronalaženje literature koja se podudara s kriterijima uključivanja i isključivanja definiranim u protokolu. Odjeljak metodologije SR-a trebao bi navesti sve baze podataka i indekse citata koji su pretraženi. Uobičajeni izvori informacija koji se koriste u pretraživanjima uključuju znanstvene baze recenziranih članaka poput MEDLINE, Web of Science, Embase i PubMed, kao i izvore neobjavljene literature kao što su registri kliničkih istraživanja i zbirke takozvane „sive literature" (32).

iii) Ekstrakcija relevantnih podataka: relevantni podaci vade se iz izvora podataka i ovise o metodi SR-a. Važno je napomenuti da je metoda ekstrakcije podataka specifična za vrstu podataka, a ekstrahirani podaci o ishodima relevantni su samo za određene vrste SR-ova (30).

iv) Procjena prihvatljivosti podataka: ova faza uključuje procjenu prihvatljivosti podataka za uključivanje u pregled, ocjenom prema kriterijima utvrđenim u prvoj fazi. Bilježi se zašto su donesene odluke o uključivanju ili isključenju iz pregleda. Za podršku procesu odabira može se koristiti i računalni program, uključujući alate za obradu teksta i strojno učenje, što može automatizirati pojedine dijelove procesa (33).

v) Analiza i kombiniranje podataka: u ovoj fazi potrebno je dobiti ukupni rezultat svih podataka. Budući da ovaj zbirni rezultat koristi kvalitativne ili kvantitativne podatke iz svih prihvatljivih izvora podataka, smatra se pouzdanijim jer pruža bolje dokaze, jer što je više podataka uključeno u preglede, možemo biti sigurniji u zaključke. Kad je prikladno, neki sustavni pregledi uključuju MA koja koristi statističke metode za kombiniranje podataka iz više izvora (27, 28, 30).

Iako se SR smatraju najjačim dokazom, pregled Mohera i suradnika iz 2003. godine pokazao je da nisu svi SR bili jednako pouzdani te da se njihovo izvješćivanje može poboljšati univerzalno dogovorenim skupom standarda i smjernica (34).

Najrigorozniji sustav izrade SR zasigurno su Cochraneovi sustavni pregledi (engl. *Cochrane systematic review – CSR*). Više pregleda SR-ova pokazalo je kako je metodologija CSR-ova bolja u odnosu na ne-Cochraneove sustavne preglede (35, 36). Osim što slijedi opisane faze izrade SR, Cochrane pri izradi svojih SR inzistira na dodatnim provjerama sigurnosti i kvalitete. Neizostavni dio metodologije CSR jesu: navodi potencijalnog sukoba interesa i procjena rizika od pristranosti (*RoB*) (37).

4. Cochrane

4.1. Organizacija Cochrane

Cochrane (ranije poznata kao *Cochrane Collaboration*) britanska je međunarodna neprofitna organizacija koja izrađuje sustavne preglede medicinskih istraživanja kako bi poboljšala i omogućila medicinsku praksu utemeljenu na dokazima o zdravstvenim intervencijama (38). Obuhvaća niz radnih skupina sa sjedištem u istraživačkim institucijama širom svijeta. Cochrane okuplja oko 30 000 volonterskih stručnjaka iz cijelog svijeta. Cochraneovi SR objavljuju u Cochrane knjižnici (engl. *Cochrane Library*), u Cochraneovoj bazi sustavnih pregleda (engl. *Cochrane Database of Systematic Reviews – CDSR*) (38, 39).

Cochrane ima viziju svijeta poboljšanog zdravlja u kojem se odluke o zdravlju i zdravstvenoj zaštiti temelje na visokokvalitetnim, relevantnim i suvremenim sintetiziranim dokazima istraživanja. Misija Cochranea je promovirati donošenje zdravstvenih odluka utemeljenih na dokazima izradom visokokvalitetnih, relevantnih, dostupnih sustavnih pregleda i ostalih sinteza znanstvenih dokaza (38).

Cochrane omogućuje pristupačne i vjerodostojne informacije kao podršku informiranom donošenju odluka. U doba Interneta ljudi imaju puno veći pristup zdravstvenim informacijama, ali malo je načina da znaju jesu li te informacije točne i nepristrane. Cochrane ne prihvaća komercijalno financirana istraživanja niti ona opterećena sukobom komercijalnih interesa. To je od vitalne važnosti za generiranje mjerodavnih i pouzdanih informacija, slobodan rad, nesputan komercijalnim i financijskim interesima (38).

4.2. Povijesni razvoj Cochranea

Tri čovjeka zaslužna su za stvaranje trenutnog Cochranea: Tom Chalmers, Iain Chalmers i Murray Enkin. Ime organizacije odaje počast Britancu Archibaldu (Archieju) Cochraneu i njegovim pionirskim naporima da izbjegne pristranost u kliničkim istraživanjima promoviranjem RCT-a i sinteza dokaza iz RCT-ova. U zarobljeništvu, tijekom Drugog svjetskog rata, Cochrane je izveo svoj prvi RCT nad ratnim zarobljenicima, uspoređujući učinak ekstrakta kvasca na bolesti različitih nedostatka kod pothranjenosti. Suosjećanje sa subjektima istraživanja, koji su bili i njegovi kolege i zarobljenici, utjecalo je na sav njegov budući rad. Načelo da RCT-ovi moraju pružiti korist subjektima obilježje je Cochranea. Cochrane je prvi značajan rad objavio dok je bio u Cardiffu u Južnom Walesu, gdje je proveo dugi niz godina pokušavajući utvrditi razliku učinka tuberkuloze naspram prašine u uzrokovanju progresivne plućne fibroze - istraživanje Rhondda Fach. Neke od najcjenjenijih lekcija koje je naučio iz

istraživanja Rhondda Fach uključivale su vrijednost epidemioloških istraživanja i rizike od pristranosti istraživanja (40). Cochrane je 1979. izrazio zabrinutost zbog nepostojanja valjane i organizirane sinteze svih RCT-a koji se odnose na svaku specijalnost ili subspecijalnost. Po njemu bi važeća sinteza svih RCT-a olakšala kliničarima donošenje točnih zaključaka (41).

Tom Chalmers proširio je Cochraneov rad. Ustvrdio je da su RCT-i temelj hijerarhije dokaza, a da su vrh hijerarhije objedinjeni podaci iz više istraživanja (SR). Dodao je da se pristranost objave (engl. *publication bias*), činjenica da će vjerojatnije biti objavljena istraživanja s pozitivnim rezultatima od onih s negativnim rezultatima, mora uzeti u obzir prilikom sažimanja i prikaza rezultata. Zaslužan je za uvođenje MA (5, 42).

U međuvremenu, porodničar Iain Chalmers tražio je učinkovite načine liječenja tijekom svojih iskustava u palestinskim izbjegličkim kampovima 1960-ih. Tradicionalno, antibiotici su indicirani samo ako postoje jasni znakovi infekcije. Kad bi roditelji doveli svoju bolesnu djecu Chalmersu, on je inzistirao da se vrate po antibiotike tek nakon što su znakovi infekcije bili prisutni. Međutim, do tada bi već bakterije bile uzrokovale ireverzibilno septično stanje kod neuhranjene, oslabljene djece. Chalmers je tako prepoznao opasnost prepuštanja medicinskoj dogmi i vitalnu važnost traženja istine (42).

Chalmers se udružio s porodničarem Murrayem Enkinom kako bi stvorio masovnu bazu podataka objavljenih, neobjavljenih, trajnih i planiranih istraživanja i MA. Baza podataka o perinatalnim istraživanjima iz Oxforda stvorila je temelj za značajnu knjigu "Učinkovita skrb u trudnoći i rođenju djeteta" (43).

Rad tih pionira doveo je do stvaranja organizacije nazvane *Cochrane Collaboration* 1993. godine. Ponavljajući način rada Oxfordove baze podataka perinatalnih istraživanja (engl. *Oxford Database of Perinatal Trials*), i danas se brojne druge medicinske specijalnosti drže deset navedenih principa prakse: suradnja, nadogradnja na entuzijazmu pojedinaca, izbjegavanje dupliciranja, umanjivanje pristranosti, ažurnost, relevantnost, promicanje pristupa, osiguravanje kvalitete, kontinuitet i međunarodna suradnja (44).

Cochrane Collaboration osnovana je 1993. godine, nakon sastanka početne skupine od 77 ljudi iz devet zemalja, na prvom *Cochrane Colloquiumu* u Oxfordu u Velikoj Britaniji. Od 2015. godine organizacija se jednostavno naziva "Cochrane". Tijekom godina organizacija je narasla na više od 13000 članova i 50000 podržavatelja iz više od 130 zemalja do 2018. godine, što je čini najvećom organizacijom uključenom u izradu i promociju SR-ova. Cochrane nastavlja širiti svoj doseg mrežama na svim kontinentima i novim grupama na razini države koje se osnivaju svake godine, a laički sažetci Cochraneovih sustavnih pregleda (engl. *plain language summary* – PLS) prevode se na 14 jezika, uključujući i hrvatski jezik (38).

5. Procjena rizika od pristranosti

5.1. Rizik od pristranosti

U kojoj mjeri SR mogu donijeti zaključke o učincima intervencije ovisi o tome jesu li podaci i rezultati uključenih istraživanja valjani. Primjerice, MA nevaljanih istraživanja može dati zavaravajući rezultat, dajući uski interval pouzdanosti oko pogrešne procjene učinka intervencije. Procjena valjanosti obuhvaćenih istraživanja stoga je bitna komponenta nekih SR, a posebno CSR i trebala bi utjecati na analizu, tumačenje i zaključke SR (45).

Može se smatrati da valjanost istraživanja ima dvije dimenzije. Prva je dimenzija postavlja li istraživanje odgovarajuće istraživačko pitanje. To se često opisuje kao „vanjska valjanost", a njena procjena ovisi o svrsi u koju će se istraživanje koristiti. Vanjska valjanost usko je povezana s generalizacijom ili primjenjivošću rezultata istraživanja (45).

Druga dimenzija valjanosti istraživanja odnosi se na to odgovara li na svoje istraživačko pitanje „ispravno", odnosno na način koji nije pristran. To se često opisuje kao „unutarnja valjanost", a upravo se na ovaj aspekt valjanosti obraćamo u ovom poglavlju. Kako se većina CSR usredotočuje na RCT, tako je glavni primjer u izradi SR procjena valjanosti te vrste istraživanja. Procjene interne valjanosti često se nazivaju „procjenama metodološke kvalitete" (45).

Pristranost je oznaka sustavne pogreške ili odstupanje od istine u rezultatima ili zaključcima. Pristranost može biti izražena u bilo kojem smjeru: može dovesti do podcjenjivanja ili precjenjivanja učinka intervencije. Jačina pristranosti može varirati: nekad je mala (i trivijalna u usporedbi s promatranim učinkom), a nekad je značajna (tako da očigledan nalaz može biti u potpunosti posljedica pristranosti). Čak se i određeni izvor pristranosti može razlikovati u smjeru: pristranost zbog nedostataka u ustroju istraživanja (npr. izostanak prikrivanja razvrstavanja – engl. *allocation concealment*) može dovesti do podcjenjivanja učinka u jednom istraživanju, ali precjenjivanja u nekom drugom istraživanju. Obično je nemoguće znati u kojoj su mjeri pristranosti utjecale na rezultate određenog istraživanja, iako postoje dobri empirijski dokazi da određeni nedostaci u ustroju, provođenju i analizi RCT dovode do pristranosti. Budući da rezultati istraživanja u stvari mogu biti nepristrani unatoč metodološkoj pogrešci, prikladnije je procjenjivati rizik od pristranosti (RoB) (46).

Razlike u rizicima od pristranosti mogu pomoći u objašnjavanju varijacija rezultata istraživanja uključenih u sustavni pregled (mogu objasniti heterogenost rezultata). Stroža istraživanja vjerojatnije će dati rezultate koji su bliži istini. MA rezultata istraživanja varijabilne valjanosti može rezultirati lažno pozitivnim zaključcima (pogrešno zaključivanje da je

intervencija učinkovita) ako su manje rigorozna istraživanja pristrana prema precjenjivanju učinka intervencije. Također bi mogli doći do lažno negativnih zaključaka (pogrešno zaključujući da nema učinka) ako su manje rigorozna istraživanja pristrana prema podcjenjivanju učinka intervencije (47).

Važno je procijeniti RoB u svim istraživanjima u SR-u, bez obzira na očekivanu varijabilnost rezultata ili valjanosti obuhvaćenih istraživanja. Na primjer, rezultati mogu biti dosljedni među istraživanjima, ali sva istraživanja mogu biti pristrana. U tom slučaju, zaključci SR-a ne bi smjeli biti toliko snažni kao da niz rigoroznih istraživanja daje dosljedne rezultate o učinku intervencije. U CSR ovaj postupak ocjenjivanja opisan je kao procjena RoB o uključenim istraživanjima (46).

Pristranost se ne smije miješati s nepreciznošću. Pristranost se odnosi na sustavnu pogrešku, što znači da bi višestruka ponavljanja istog istraživanja u prosjeku dovela do sličnog pogrešnog rezultata. Nepreciznost se odnosi na slučajnu pogrešku, što znači da će višestruke replikacije istog istraživanja proizvesti različite procjene učinka zbog varijacije uzorkovanja, čak i ako bi u prosjeku dala ispravan rezultat. Rezultati manjih istraživanja podložni su većim varijacijama uzorkovanja i stoga su manje precizni. Nepreciznost se odražava u intervalu pouzdanosti oko procjene intervencijskog učinka iz svakog istraživanja i u težini koja se pridaje rezultatima svakog istraživanja u MA. Precizniji rezultati daju veću težinu (46).

Pristranost treba razlikovati od kvalitete. Izraz "procjena metodološke kvalitete" uvriježeno se koristi u kontekstu metoda SR-a i označava kritičku ocjenu obuhvaćenih istraživanja. Izraz sugerira procjenu u kojoj su mjeri autori provodili svoja istraživanja prema najvišim mogućim standardima. Cochraneov priručnik pravi razliku između procjene metodološke kvalitete i procjene RoB i preporučuje usredotočenje na potonju (46).

Ključno razmatranje u CSR je mjera u kojoj treba vjerovati rezultatima obuhvaćenih istraživanja. Procjenom RoB to je pitanje ciljano. Istraživanje se može izvesti prema najvišim mogućim standardima, ali i dalje može imati važan RoB. Na primjer, u mnogim je situacijama nepraktično ili nemoguće zaslijepiti (engl. *blinding; masking*) ispitanike ili osoblje istraživanja u smislu da ne znaju tko je u intervencijskoj skupini. Neprimjereno je opisati sva takva istraživanja kao „nekvalitetna", ali to ne znači da nemaju pristranosti koje proizlaze iz znanja o statusu intervencije (34).

Neki pokazatelji kvalitete u medicinskim istraživanjima, kao što je dobivanje etičkog odobrenja, izračun veličine uzorka ili izvješćivanje o istraživanju u skladu s izjavom CONSORT (48) neće imati izravne implikacije na RoB. Naglasak na RoB prevladava dvosmislenost između kvalitete izvješćivanja i kvalitete temeljnog istraživanja (iako ne

prevladava problem potrebe da se oslanja na izvješća za procjenu temeljnog istraživanja). Bez obzira na ograničenja, RoB u rezultatima svakog istraživanja koje doprinose procjeni učinka jedan je od nekoliko čimbenika koji se moraju uzeti u obzir prilikom ocjenjivanja kakvoće dokaza (34).

Pristranost povezana s određenim karakteristikama istraživanja može se ispitati uporabom metode koja je poznata kao meta-epidemiologija (49, 50). Meta-epidemiološko istraživanje analizira skupinu MA, u kojima je svako uključeno istraživanje klasificirano prema nekim karakteristikama na razini istraživanja. Rani primjer bilo je proučavanje kliničkih istraživanja s dihotomnim ishodima uključenim u MA iz CSR o trudnoći i porođaju (51). To je istraživanje pokazalo da su istraživanja u kojima je randomizacija bila neadekvatno prikrivena ili o njoj nije bilo izvješćivano, dala pretjerane procjene učinka intervencije u usporedbi s pokusima koji su prijavili odgovarajuće prikrivanje, te je pronašla sličnu (ali manju) povezanost za istraživanja koja nisu opisana kao dvostruko slijepa (engl. *double-blind*) (49, 50).

5.2. Cochraneov alat za procjenu rizika od pristranosti iz 2011. godine

Procjena valjanosti RCT-ova uključenih u Cochraneove sustavne preglede temelji se na alatu koji se naziva Cochraneov alat za procjenu rizika od pristranosti (engl. Cochrane risk of bias tool) (52). Tim se rizikom opisuje precijenjenost ili podcijenjenost pravog učinka istraživane intervencije (47, 53-55).

Autori Cochraneovih sustavnih pregleda trebali bi prema tih sedam domena procijeniti rizik od pristranosti za svaki RCT uključen u sustavni pregled. Rizik za svaku domenu procjenjuje se kao visok, nizak ili nejasan. Kategorija nejasnog rizika upućuje na nedostatnost informacija ili nesigurnost u razinu potencijalnog rizika. Za svaku domenu, osim procjene rizika, Cochraneovi autori trebaju navesti i popratni komentar u kojem obrazlažu razloge za svoju procjenu. Odluke donesene prema Cochraneovom alatu za rizik od pristranosti bilježe se u standardiziranu tablicu rizika sustavne pogreške koja za svaki RCT sadrži ime RCT-a, i sedam redova u kojima se nalaze po tri polja: naslov domene, procjena rizika i objašnjenje procjene (52).

Cochraneov alat za procjenu rizika od pristranosti u RCT-ovima ima sedam domena (46):

i) Generiranje nasumičnog redoslijeda (engl. *random sequence generation; selection bias*) je domena u kojoj se procjenjuje postojanje moguće pristranosti zbog neodgovarajućeg načina randomiziranja ispitanika.

   Upotreba nasumične komponente trebala bi biti dovoljna za prikladno generiranje redoslijeda i kao takva procijenjena niskim RoB za ovu domenu. Randomizacija bez

ograničenja za generiranje niza raspodjele naziva se jednostavna randomizacija. Primjerice, to se može postići dodjeljivanjem intervencija metodama poput ponovljenog bacanja novčića, bacanja kockica ili dijeljenja prethodno izmiješanih karata (56, 57). Obično se to postiže pozivanjem na objavljeni popis nasumičnih brojeva ili na popis nasumičnih zadataka koje generira računalo. U pokusima koji koriste velike uzorke (obično to znači najmanje 100 u svakoj randomiziranoj skupini) (56-58), jednostavnom randomizacijom generiraju se usporedne skupine relativno sličnih veličina. U pokusima s malim uzorcima jednostavna randomizacija ponekad će rezultirati dodjelom redoslijeda koji dovodi do skupina koje se slučajno prilično značajno razlikuju u veličini ili u pojavi prognostičkih čimbenika (59).

Kao procjene s visokim RoB za ovu domenu često se koriste neprikladne metode, poput izmjenjivanja (alternacije), dodjele na temelju datuma rođenja, broja evidencije slučaja i datuma prezentacije, a ponekad se takve metode nazivaju i kvazi-nasumičnim (engl. *quasi-random*). Izmjena (ili rotacija, za više od dvije intervencijske skupine) u principu može rezultirati sličnim skupinama, ali mnoge druge sustavne metode generiranja redoslijeda ne moraju. Primjerice, dan kad je pacijent primljen u bolnicu nije samo stvar slučaja (60).

Jednostavna izjava poput "nasumično smo dodijelili" ili "istraživanje je bilo randomiziranog ustroja" često je nedovoljna da bismo bili sigurni da je redoslijed raspodjele ispitanika po skupinama stvarno randomiziran. Nerijetko se događa da autori koriste izraz "randomizirano" čak i kada to nije opravdano: mnogi pokusi s deklariranom kvazi-nasumičnom raspodjelom autori opisuju kao randomizirana. Ako postoji sumnja, tada se procjena RoB za generiranje nasumičnog redoslijeda treba smatrati nejasnom (engl. *unclear*) (60).

Ponekad autori istraživanja daju neke informacije, ali oni u potpunosti ne definiraju svoj pristup i ne potvrđuju neke nasumične komponente u procesu. Na primjer, autori mogu navesti da je korištena randomizacija u bloku, ali postupak odabira blokova, poput tablice nasumičnih brojeva ili računalnog generatora nasumičnih brojeva, nije naveden. Procjenu generiranja redoslijeda tada treba klasificirati kao nejasnu (60).

ii) Prikrivanje razvrstavanja (engl. *allocation concealment; selection bias*) je domena u kojoj se procjenjuje moguća pristranost zbog neodgovarajućeg načina sakrivanja nasumičnog redoslijeda. Ispravno prikrivanje razvrstavanja osigurava strogu provedbu redoslijeda bez predznanja intervencijskih zadataka. Metode prikrivanja redoslijeda razvrstavanja odnose se na tehnike korištene za provođenje redoslijeda, a ne za njegovo

generiranje. Nije rijetkost da se dobro generiran nasumični redoslijed neadekvatno sakrije, na primjer ako je redoslijed objavljen na zidu sobe za osoblje (51).

Neki autori SR miješaju prikrivanje razvrstavanja sa zasljepljivanjem o dodijeljenim intervencijama. Prikrivanje razvrstavanja nastoji spriječiti pristranost odabira u dodjeli intervencije zaštitom redoslijeda raspodjele prije i do dodjele, a uvijek se može uspješno provesti bez obzira na temu istraživanja (51, 55). Suprotno tome, zasljepljivanjem se želi spriječiti pristranost izvedbe i otkrivanja zaštitom niza nakon dodjele i ne može se uvijek primijeniti - na primjer, u pokusima u kojima se uspoređuje kirurški i ne-kirurški medicinski tretman. Dakle, prikrivanje razvrstavanja do točke dodjele intervencije i zasljepljivanje nakon te točke rješavaju različite izvore pristranosti i razlikuju se u njihovoj izvedivosti (55, 61).

Važnost prikrivanja razvrstavanja može ovisiti o mjeri u kojoj potencijalni ispitanici istraživanja imaju različite prognoze, postoje li snažna uvjerenja među istražiteljima i ispitanicima u pogledu koristi ili štete dodijeljenih intervencija i prihvaćaju li neizvjesnost u vezi s intervencijama svi uključeni (62). Među različitim metodama koje se koriste za prikrivanje razvrstavanja, možda je najpoželjnija središnja randomizacija od strane treće osobe. Metode korištenja omotnica podložnije su manipulaciji od ostalih pristupa. Ako istražitelji koriste omotnice, trebali bi razviti i nadzirati postupak dodjele kako bi sačuvali prikrivanje. Uz upotrebu uzastopno numeriranih, neprozirnih, zatvorenih omotnica (engl. *sequentially numbered, opaque, sealed envelopes – SNOSE*), trebali bi osigurati da se omotnice otvaraju uzastopno i to tek nakon što je omotnica nepovratno dodijeljena ispitaniku (51).

iii) Zasljepljivanje ispitanika i osoblja (engl. *blinding of participants and personnel; performance bias*) je domena o procjeni pristranosti mogućnosti da ispitanici i osoblje znaju koju intervenciju ispitanici primaju.

U kliničkom istraživanju može biti zaslijepljeno više skupina subjekata: ispitanici, zdravstveno osoblje koje pruža skrb, osobe koje bilježe ishode, analitičari podataka (statističari) i osobe koje pišu tekst istraživanja (63, 64).

Nedostatak zasljepljivanja ispitanika ili osoblja mogao bi utjecati na stvarne ishode ispitanika u istraživanju. To može biti zbog izostanka očekivanog djelovanja intervencije u kontrolnoj skupini ili zbog različitog ponašanja među skupinama ispitanika (primjerice, različit broj napuštanja istraživanja, različit broj prijelaza na alternativnu intervenciju ili različito provođenje ko-intervencija) (65-68).

U empirijskim istraživanjima pokazalo se da je nedostatak zasljepljivanja u RCT-ovima povezan s pretjeranim procjenama učinaka intervencija (69). Ta su se istraživanja bavila različitim ishodima, od kojih su neki bili i objektivni. Primijećeno je da je procijenjeni učinak u prosjeku pristraniji u pokusima s više subjektivnih ishoda (70).

Zasljepljivanje može biti nemoguće za barem neke subjekte (npr. većina pacijenata koji se operiraju). Međutim, u takvim istraživanja mogu se poduzeti i druge mjere za smanjenje RoB, poput liječenja pacijenata prema strogom protokolu kako bi se smanjio rizik od različitog ponašanja pacijenata i pružatelja zdravstvenih usluga. Pokušaj zasljepljivanja ispitanika i osoblja ne osigurava uspješno zasljepljivanje u praksi, stoga je ono redovito ugroženo za većinu intervencija. Za mnoga zaslijepljena istraživanja lijekova, nuspojave lijekova omogućuju moguće otkrivanje koja se intervencija primjenjuje na koje ispitanike, osim ako istraživanje uspoređuje dvije prilično slične intervencije, npr. lijekove sa sličnim nuspojavama ili koristi aktivni placebo (65-68, 71).

Izvješća o istraživanjima često opisuju zasljepljivanje općenitim pojmovima, kao što je "dvostruko slijepo" (engl. *double-blind).* Taj pojam ne daje informaciju tko je zaslijepljen (61). Takvi se izrazi također koriste vrlo nedosljedno (64, 71, 72), a učestalost izričitog izvješćivanja o statusu zasljepljivanja ispitanika i osoblja i dalje je niska čak i u istraživanjima objavljenim u vrhunskim časopisima, unatoč preporukama CONSORT-a da mora biti izričita (48).

Kada se razmatra RoB zbog nedostatka zasljepljivanja ispitanika i osoblja, važno je posebno razmotriti dva pitanja (60):

1. Tko jest, a tko nije bio zaslijepljen
2. Postoji li rizik pristranosti u stvarnim ishodima zbog nedostatka zasljepljenja tijekom istraživanja (npr. zbog suradnje ili različitog ponašanja)

RoB može biti velik za neke ishode, a nizak za druge, čak i ako su isti ljudi nisu bili zaslijepljeni u istraživanju. Na primjer, znanje o dodijeljenoj intervenciji može utjecati na ishode u ponašanju (poput broja posjeta klinici), dok neće utjecati na ukupnu smrtnost. Prema tome, procjenu RoB koji proizlaze iz nedostatka zasljepljivanja ponekad je potrebno uraditi odvojeno za različite ishode (60).

Umjesto da se procjenjuje RoB za svaki ishod zasebno, često je prikladno grupirati ishode sa sličnim rizicima pristranosti. Na primjer, može se uraditi zajednička procjena RoB za sve subjektivne ishode koja se razlikuje od zajedničke procjene zasljepljivanja za sve objektivne ishode (60).

iv) zasljepljivanje osoba koje bilježe ishode (engl. *blinding of outcome assessors; detection bias*) je domena o procjeni RoB ako je moguće da osobe koje analiziraju ishode znaju koju intervenciju ispitanici primaju. Ako su osobe koje bilježe ishode svjesne dodjele intervencije ispitanicima, te bi procjene ishoda mogle biti pristrane. Procjenu ishoda mogu izvršiti sami ispitanici, njihovi pružatelji zdravstvenih usluga ili neovisni procjenitelji (60).

Na sve procjene ishoda može utjecati nedostatak zasljepljivanja, iako postoje posebni rizici od pristranosti sa subjektivnijim ishodima (npr. bol ili broj dana s prehladom). Stoga je važno razmotriti koliko je ishod subjektivan ili objektivan kad se razmišlja o zasljepljivanju. Važnost zasljepljivanja i mogućnost zasljepljivanja mogu se razlikovati po rezultatima unutar istraživanja. Naizgled objektivne procjene, primjerice povezane s liječnicima koji procjenjuju stupanj psihološkog ili tjelesnog oštećenja, sigurno su dobrim dijelom subjektivne (73).

Zasljepljivanje procjene ishoda može biti nemoguće (npr. kada su pacijenti podvrgnuti operaciji). Međutim, to ne znači da se potencijalne pristranosti mogu zanemariti, a autori sustavnih pregleda i dalje bi trebali procijeniti RoB visokim rizikom zbog nedostatka zasljepljivanja procjene ishoda za sva istraživanja u njihovom pregledu (60).

v) Nepotpuni podatci o ishodima (engl. *incomplete outcome data; attrition bias*) je domena o procjeni pristranosti zbog gubitka ispitanika iz istraživanja i posljedičnih nepotpunih podataka o randomiziranim ispitanicima. Podaci o ishodima koji nedostaju zbog gubitka ispitanika (odustajanja) tijekom istraživanja ili isključenja iz analize, povećavaju mogućnost da je promatrana procjena učinka pristrana. Koristi se izraz nepotpuni podaci o ishodima kako bi se odnosili i na osipanje i isključenja, a nestalim ćemo smatrati ishod pojedinog ispitanika koji nije dostupan (74).

Gubitak ispitanika se može pojaviti iz sljedećih razloga: ispitanici se povlače ili su povučeni iz istraživanja; ispitanici ne prisustvuju kontrolnom pregledu na kojem je trebalo mjeriti ishode; ispitanici prisustvuju kontroli, ali ne daju relevantne podatke; ispitanici ne popunjavaju dnevnike ili upitnike; ispitanici se ne mogu locirati (izgubljeni tijekom praćenja – engl. *lost to follow-up*); istražitelji odlučuju, obično neprimjereno, prekinuti praćenje; podaci ili zapisi su izgubljeni ili su nedostupni iz drugih razloga. Uz to, neki ispitanici mogu biti isključeni iz analize iz sljedećih razloga: neki su upisani ispitanici kasnije utvrđeni kao neprihvatljivi; izvodi se analiza „prema tretmanu“ (ili prema protokolu) (u koju su ispitanici uključeni samo ako su primili namjeravanu

intervenciju u skladu s protokolom); analiza istraživanja isključila je neke ispitanike iz drugih razloga (75).

vi) Selektivni prikaz rezultata (engl. *selective reporting; reporting bias*) domena je o procjeni pristranosti zbog mogućeg neobjavljivanja svih rezultata koji su dobiveni istraživanjem. Selektivno izvješćivanje o ishodu definirano je kao odabir podskupine izvornih varijabli zabilježenih, na temelju rezultata, radi uključivanja u publikaciju istraživanja (74).

Selektivno izvješćivanje o rezultatima može nastati na nekoliko načina: selektivni prikaz ishoda; selektivni odabir podataka za ishod; selektivno izvješćivanje različitih analiza koje su koristile iste podatke; selektivno izvješćivanje o podskupovima podataka; selektivno kratko izvješćivanje o podacima (vezano direktno za statistički rezultat) (76).

vii) Ostalo (engl. *other bias*) je domena koja pokriva sve ostale moguće izvore pristranosti koji nisu pokriveni s prvih šest domena (46).

5.3. RoB 2 alat

Nakon gotovo desetljeća iskustva korištenja Cochraneovog RoB 2011 alata, uočena su potencijalna poboljšanja na temelju pregleda literature, iskustva korisnika i povratnih informacija, pristupa korištenih u drugim alatima za rizik od pristranosti i novijeg razvoja metodologije u procjeni intervencijskih učinaka iz RCT-a. Zbog toga je razvijen revidirani alat za procjenu rizika od pristranosti u randomiziranim istraživanjima (RoB 2) (77).

RoB 2 procjenjuje pristranost u pet različitih domena. Unutar svake domene korisnici RoB-a 2 moraju odgovoriti na jedno ili više tzv. signalnih pitanja (engl. *signaling questions*). Ti odgovori vode do procjene "nizak rizik od pristranosti", "visoki rizik od pristranosti" ili, za razliku od prethodnog alata, više se ne koristi izraz „nejasan rizik od pristranosti“, već "određene zabrinutosti" (engl. *some concerns*).Usto, procjene unutar svake domene dovode do izračuna ukupne procjene o RoB za svaki promatrani rezultat/ishod, što bi trebalo omogućiti korisnicima RoB 2 da stratificiraju meta-analize prema ukupnom RoB (78).

RoB 2 uveden je u praksu izrade CSR nakon što je objavljen sredinom 2019. godine, nakon što su bile obavljene sve analize iz ove disertacije. Još uvijek nije u potpunosti implementiran u Cochraneovim sustavnim pregledima te će sigurno u budućnosti zahtijevati nove promjene koje se mogu temeljiti i na rezultatima radova opisanih u ovoj disertaciji, i srodnih radova (78-88).

CILJEVI I HIPOTEZE

Ciljevi

Glavni cilj ove disertacije je analizirati učestalost različitih procjena RoB kroz prve četiri domene Cochraneovog RoB iz 2011. u RCT-ovima s intervencijama kirurškog tipa. Drugi glavni cilj je analiza podudarnosti procjena Cochraneovih autora s našim procjenama RoB u RCT-ovima s intervencijama kirurškog tipa kroz ove četiri odabrane domene koje su jasnije definirane u Cochraneovom priručniku.

Sporedni cilj je analiza i usporedba učestalosti procjena RoB od strane Cochraneovih autora, a i naših procjena RoB između RCT-ova s kirurškim intervencijama i onih s ne-kirurškim intervencijama. Još jedan sporedni cilj je bilo otkrivanje mogućih generaliziranih pogreški u procesima procjene.

Hipoteze

U Cochraneovim sustavnim pregledima:

1. U domenama generiranja nasumičnog redoslijeda i prikrivanja razvrstavanja (prva i druga domena Cochraneovog alata za procjenu RoB) učestalost nejasne procjene rizika od pristranosti (engl. *unclear RoB*) u RCT-ovima s intervencijama kirurškog tipa jednaka je kao i u RCT-ovima o ne-kirurškim intervencijama.
2. U domenama o zasljepljivanju ispitanika i osoblja te zasljepljivanju osoba koje bilježe ishode (treća i četvrta domena Cochraneovog alata za procjenu RoB) učestalost nejasne procjene rizika od pristranosti (engl. *unclear RoB*) u RCT-ovima s intervencijom kirurškog tipa manja je nego u RCT-ovima o ne-kirurškim intervencijama.
3. Učestalost nepodudarnih procjena Cochraneovih autora s našim procjenama RoB veća je u RCT-ovima s intervencijama kirurškog tipa nego u RCT-ovima o ne-kirurškim intervencijama kroz sve četiri prve domene Cochraneovog alata za procjenu RoB ukupno i pojedinačno.

Naglasak u našim istraživanjima je, stoga, na neusklađenost u odnosu na upute iz Cochraneovog priručnika, a ne na apsolutnu istinu. Naše procjene smo stoga imenovali kao podudarne ili nepodudarne s procjenama Cochraneovih autora, kako bismo naglasili relativnost usporedbe.

METODE

1. Uzorak i zajedničko prikupljanje podataka

1.1. Uzorak analiziranih Cochraneovih sustavnih pregleda

Analizirani su svi CSR koji su uključivali RCT-ove, i koji su objavljeni u CDSR od početka srpnja 2015. do kraja lipnja 2016. godine, uz korištenje naprednih opcija pretraživanja. Radi se o konvencionalnom uzorku značajne veličine (N = 955 CSR) u vremenu kad je upotreba Cochraneovog alata za procjenu rizika pristranosti iz 2011 bila obavezan dio metodologije CSR-ova. Iz promatranja su isključeni CSR-ovi (N = 237) koji nisu sadržavali RCT-ove o intervencijama (dijagnostički CSR-ovi, pregledi sustavnih pregleda itd.) kao i prazni pregledi, pregledi povučeni iz Cochraneove baze tijekom analiziranog razdoblja te svi CSR-ovi koji su sadržavali samo nerandomizirana istraživanja. Ako je sustavni pregled uključivao RCT-ove i nerandomizirana istraživanja analizirane su tablice procjene rizika od pristranosti (RoB tablica) za sva istraživanja o RCT-ovima unutar CSR-a. Jedan autor je analizirao sve naslove i sažetke kako bi utvrdio kriterije uključenja, a potom je drugi autor potvrdio procjenu prvog autora.

1.2. Prikupljanje podataka

Iz svakog CSR-a izvađeni su sljedeći podatci:

- vrsta uključenih istraživanja
- broj uključenih RCT-ova
- detaljni podatci svake RoB tablice: ime RCT-a, imena RoB domena, procjena RoB-a za svaku domenu, i popratni komentar

Prikupljanje (engl. *data scraping*) podataka iz svake RoB tablice je bilo automatizirano kroz Microsoft Excel 2010 (Microsoft, Redmond, WA, SAD) te korištenje makro-naredbi kodiranih u programskom alatu *Visual Basic for Applications* (*VBA*, Microsoft, Redmond, WA, SAD). Računalno programiranje napravio je pristupnik Ognjen Barčot. Prvo se uradilo kopiranje cjelokupnog sadržaja svakog pojedinog CSR-a, uključenog u analizu, u zasebne radne listove MS Excel knjige sa internetskih stranica Cochraneove knjižnice, odnosno baze CDSR. Potom je iz velike količine neobrađenih podataka bilo nužno izdvojiti RoB tablice za svako istraživanje unutar svakog CSR-a. Taj je proces također bio automatiziran zasebnim makronaredbama. To je bilo moguće zbog jednoobrazne organizacije internetskih stranica unutar CDSR za svaki CSR. Naime, mrežna stranica svakog CSR-a sadrži odjeljak koji se zove „Obilježja uključenih

istraživanja“ (engl. *Characteristics of included studies*) u kojima se za svako uključeno istraživanje najprije navede ime istraživanja (jedinstveni identifikator koji je najčešće naveden po modelu „prezime prvog autora/zarez/godina objave rada), zatim tzv. PICO tablica (od engl. *Patients, Intervention, Comparison, Outcomes*) koja opisuje obilježja ispitanika, intervencija, kontrole i ishoda, a nakon te tablice slijedi RoB tablica koja ima naslov „Rizik od pristranosti“ (engl. *Risk of Bias*).

Pretraga je urađena tako da su se u Google tražilici pretraživali naslovi Cochraneovih sustavnih pregleda dobivenih u obliku teksta (liste) izvezene nakon naprednog pretraživanja na stranicama Cochraneove knjižnice, potom je u tekstu bio identificiran alfanumerički niz (engl. *string*) „onlinelibrary.wiley.com/doi/“ koji je (kao dio tražene Internet adrese (engl. *URL - Uniform Resource Locator*)), potom bio kopiran u listu adresa. Iz sadašnje perspektive kada se jednostavnom pretragom URL-ovi mogu izvesti (engl. *export*) kroz sučelje napredne pretrage ranija ekstrakcija adresa je složenija, ali u potpunosti valjana. S obzirom na vremenski odmak pretraga se ne može ponoviti (do sada su nastale nove verzije – pub# koje se ne mogu otvoriti), ali DOI ID (8 alfanumeričkih znakova) se podudaraju u svim slučajevima kao i u ranijoj pretrazi što u potpunosti potvrđuje adrese ranije pretrage.

Radi transparentnosti i provjerljivosti, otvoren je projekt na mrežnoj stranici Open Science Framework, na linku https://osf.io/fmjxz/, na kojem su postavljene posljednje verzije svih ekstrakcija podataka, kako bi se mogla provjeriti točnost ekstrakcija u usporedbi s informacijama koje se nalaze u objavljenim Cochraneovim sustavnim pregledima. Valja napomenuti da je na OSF postavljen sadržaj koji odgovara originalnim tablicama koje su korištene u ovom radu u svojem necenzuriranom obliku.

Kada se postiglo da se posljednje verzije svih CSR-ova kopiraju u listove MS Excel tablice razvijen je program za ekstrakciju samih RoB tablica iz ukupno kopiranog sadržaja. Većina podataka je bila uredno ekstrahirana i u uređenoj finalnoj tablici (Nakon_ekstrakcije_v40_s_linkovima.xlsm – Sheet 2; dostupno na stranici https://osf.io/fmjxz/) je bilo razvidno da svi dijelovi nisu prisutni. Tada je bilo nužno revidirati programski kod, uvesti dodatne makronaredbe kako bi se ekstrakcija dovela do potpunosti. Zbog diversifikacije vidljive u posljednjim stupcima (Nakon_ekstrakcije_v40_s_linkovima.xlsm – Sheet 2 - stupci AA do DJ) raspodjela je morala biti urađena ručno unutar tablice.

Unatoč dosljednoj organizaciji mrežnih stranica pojedinih CSR-ova, proces automatskog prikupljanja podataka zahtijevao je niz koraka u programiranju, traženje pogreški i testiranje kako bi se izbjegle pogreške i označila mjesta gdje nedostaju ključni podatci. Programski kod

pripremljen za potrebe ovog istraživanja mogu se upotrijebiti samo za ovu specifičnu svrhu i ne mogu se koristiti za prikupljanje nikakvih drugih podataka bez reprogramiranja. Inicijalno testiranje programskog koda otkrilo je potrebu za dodatnim prilagodbama. Svaka nedosljednost podataka koji su bili obrađivani uzrokovala je grešku u kodu te njegov prekid ili upis koda pogreške u predviđeno polje u tablicama. Stoga je bilo nužno višekratno testirati i poboljšavati verzije računalnog programa za ekstrakciju podataka kako bi se otklonile sve pogreške. Ovaj dio testiranja prilikom izrade softverskih rješenja naziva se alfa-testiranjem i provođeno je sve dok nije utvrđeno da je pokrivena većina scenarija koji se mogu naći u RoB tablicama (npr. neke RoB tablice ne sadrže informacije za sve domene). Kako se ipak radilo o inačici programa koji koristi samo subjekt koji ga razvija manji dio ekstrakcije (s utvrđenim nedosljednostima podataka) je bilo lakše obaviti i dopuniti ručno. Uočeno je da u brojnim RoB tablicama nedostaju određene standardne domene, ali ne zato jer je programiranje krivo postavljeno nego jer doista nedostaju u objavljenim CSR-ovima. Također je uočeno da Cochraneovi autori često preimenuju standardne domene i uvode dodatne nove domene i nove pod-domene. Za neke uključene RCT-ove potpuno su nedostajale RoB tablice. Stoga su u programski kod ugrađena rješenja koja omogućuju prepoznavanje standardnih domena koje nedostaju, a isto tako i rješenja za izdvajanje nestandardnih domena pomoću novih algoritama. Svaka prepoznata pogreška ovog tipa upisivala bi u polja tablice kod pogreške. Kako bi se analizirale sve navedene specifičnosti, RoB tablice su izvađene zasebno, u jednakom izdanju kako su bile objavljene u CDSR, i naknadno su provjerene i uređene ručno. Domene koje su nedostajale kao i nepostojeće RoB tablice su zabilježene dijelovima koda koji upravljaju kad se pojavi pogreška te je za iste provedena naknadna ručna analiza.

Kompletiranje i povezivanje podataka napravljena je na sljedeći način:

1. Svi podatci iz zadnjih redaka svakog radnog lista su kopirani u zajednički radni list (Sheet 2).
2. Svako istraživanje koje je imalo navod koji završava znakom uskličnika (svaki kod pogreške završavao je upisom ovog znaka, na primjer: Nema domene!! ili Nema RoB!!) je provjerena ručno.
3. Pritom je za svako ovakvo pojedino istraživanje urađena provjera postoje li standardne domene među domenama s ne-standardnim naslovima te je ista bila ručno prepisana u pripadajuće polje.

2. Metode rada o prikladnosti procjena Cochraneovih autora za domenu rizika pristranosti o generiranju nasumičnog redoslijeda

2.1. Kalibracija kategorizacije

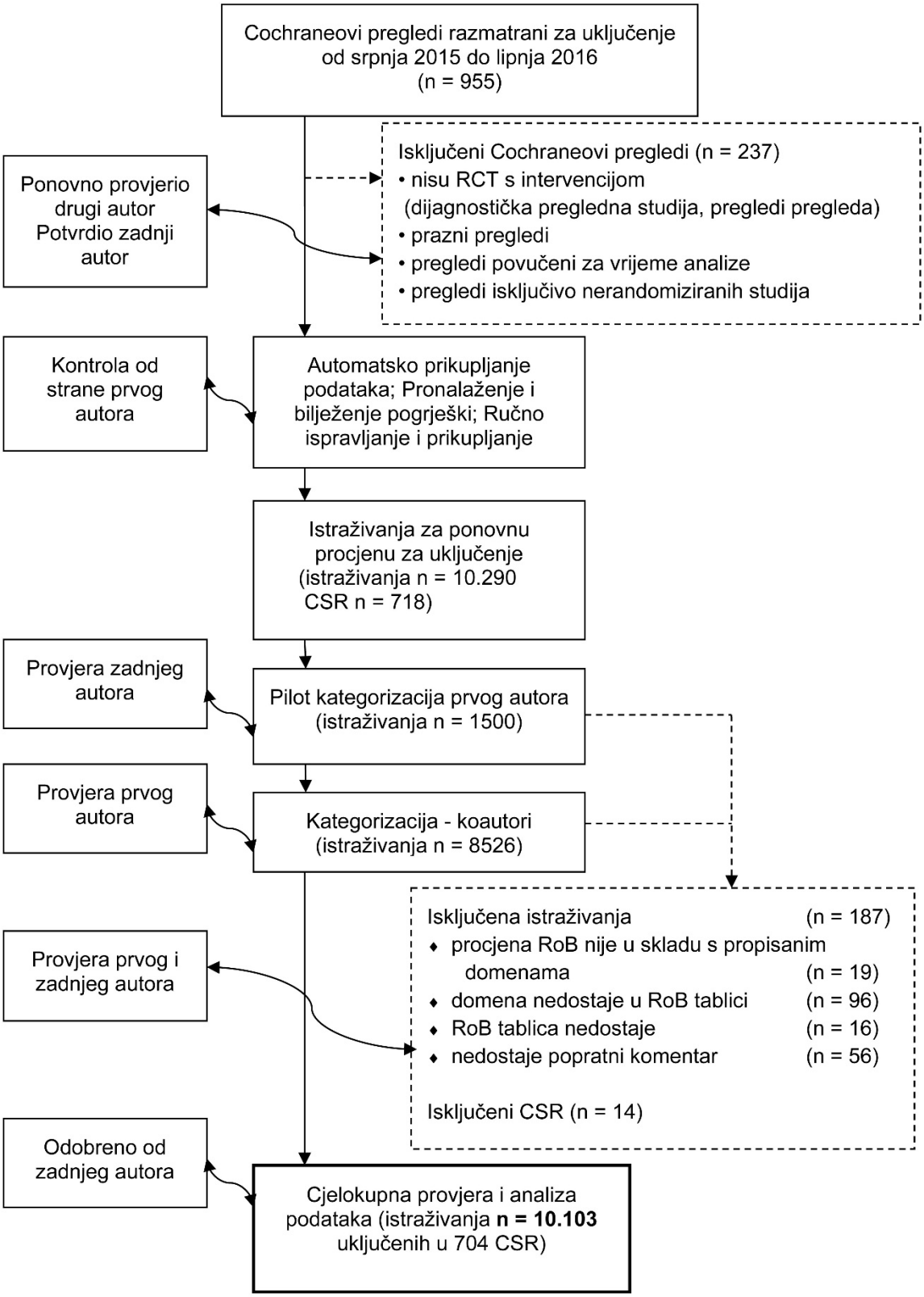
Doktorand je analizirao prvih 1500 istraživanja i svrstao ih u kategorije navedene u Cochraneovom priručniku, što je verificirala mentorica (Slika 1). Ova vježba kalibracije korištena je za izradu tablice s padajućim izbornicima koji su uključivali unaprijed određene kategorije. Daljnje su kategorizacije provodili koautori, a zatim ih je verificirao doktorand. Svi autori uključeni u kategorizacije medicinski su stručnjaci upoznati s Cochraneovim RoB alatom.

2.2. Ishodi

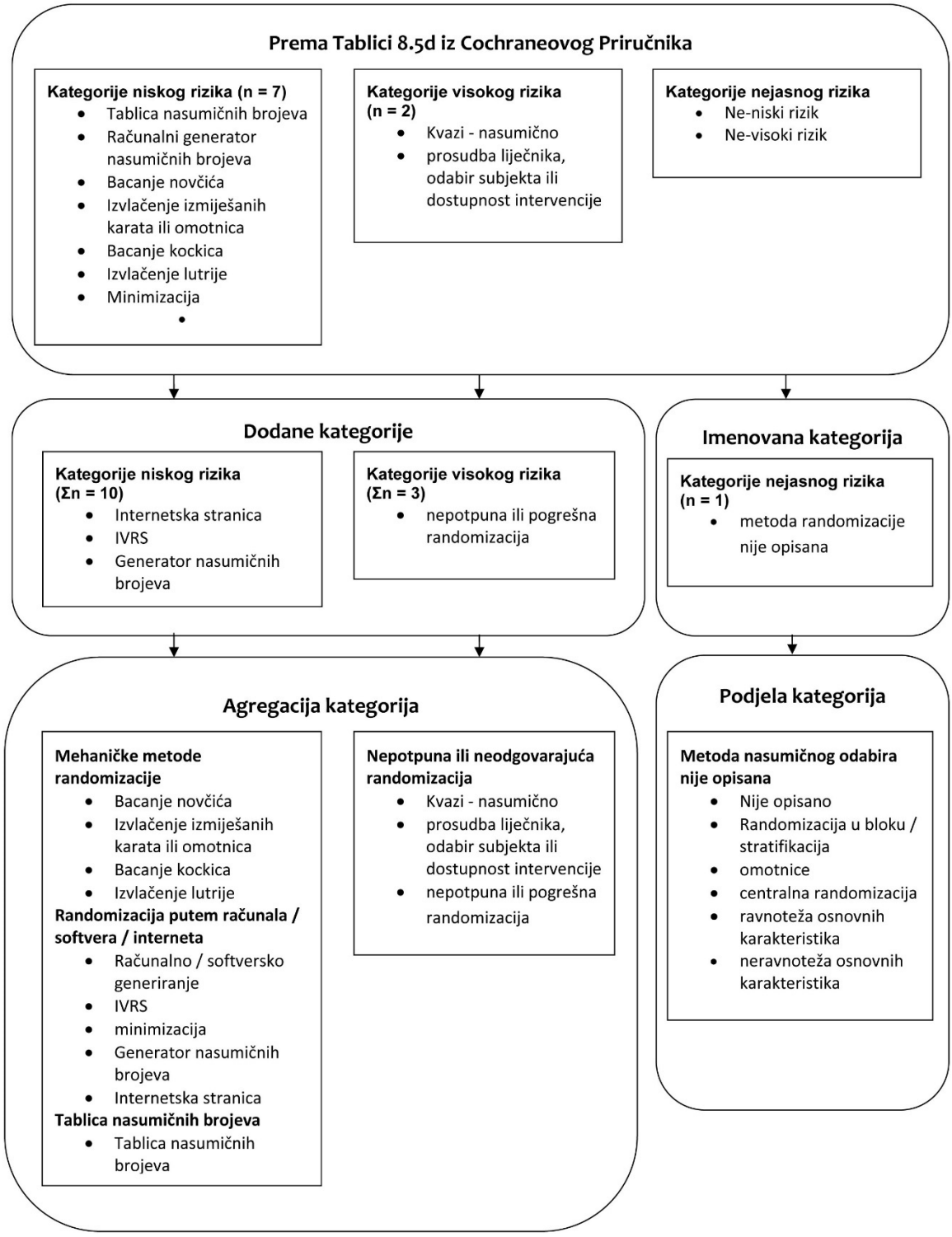
Analizirali smo broj i vrstu različitih pratećih komentara za Cochraneovu RoB domenu generiranja nasumičnog redoslijeda. Također smo analizirali podudarnost procjena za ovu domenu, u skladu s preporukama iz Cochraneova priručnika. Priručnik je korišten kao zlatni standard u našoj procjeni; ako su procjene autora Cochrane bile u suprotnosti sa smjernicama iz Cochrane priručnika, smatrali smo ih pogrešnima tj. nepodudarnima.

2.3. Kategorizacije

Kako bismo kategorizirali prateće komentare, koristili smo se preporukama iz tablice 8.5.d Cochraneovog priručnika koja sažima kriterije za procjenu RoB. Sve prateće komentare, koje su Cochraneovi autori naveli kako bi objasnili svoju procjenu, kategorizirali smo u kategorije (Slika 2).



Slika 1 - Dijagram tijeka uključivanja i isključivanja istraživanja



Slika 2 - Kategorizacija, pod-kategorizacija i agregacija

Nizak rizik od pristranosti

Kao primjere niskog RoB koristili smo svih sedam primjera iz ovog alata kao neovisne kategorije koje podržavaju procjenu (tablica nasumičnih brojeva, računalni generator nasumičnih brojeva, bacanje novčića, izvlačenje promiješanih karata ili omotnica, bacanje kockica, izvlačenje lutrije i minimizacija) . Uz to, stvorili smo zasebnu kategoriju za komentare koji su navodili da se kao metoda randomizacije koristila internetska stranica (tamo gdje je bila navedena internetska adresa) te još jedna kategorija za Interaktivni Sustav Glasovnog Odgovora (engl. *Interactive Voice Response System - IVRS*). Ako su autori Cochranea samo naveli da je korišten "generator nasumičnih brojeva", bez spominjanja računala, to je svrstano u zasebnu kategoriju - "generator nasumičnih brojeva (bez spomena računala)".

Visok rizik od pristranosti

Prema Cochraneovom priručniku, dvije kategorije podržavaju procjenu visokog rizika: kvazi-randomizacija za svaku vrstu randomizacije koja koristi neki oblik sustavne ne-nasumične komponente (poput izmjene po datumima) i randomiziranje prema prosudbi liječnika, odabiru subjekta ili dostupnosti intervencije. Dodali smo još jednu kategoriju procjene visokog rizika - nepotpuna ili pogrešna randomizacija kad su komentari ukazivali da je randomizacija djelomična ili na bilo koji način manjkava.

Nejasan rizik od pristranosti

Za svaki komentar koji ne dopušta kategorizaciju ni u jednu od 13 kategorija komentara s visokim ili niskim rizikom stvorili smo novu kategoriju - metoda nasumičnog odabira nije opisana. To je uključivalo sve slučajeve kada su autori Cochranea naznačili da nešto nije opisano ili su samo naznačili da je istraživanje opisano kao randomizirano, bez spominjanja metode randomizacije. “Metoda nasumičnog odabira nije opisana” korištena je i kada smo u popratnim komentarima pronašli navode kao što su: *nema podataka, nema dostupnih podataka, nije opisano, nije navedeno, nije prijavljeno* i slično. Slično ovome, “metoda nasumičnog odabira nije opisana” korištena je i ako su autori CSR napisali samo da se podaci o randomizaciji nalaze na određenoj stranici / tablici / slici, ali bez ikakvih detalja što je tamo napisano; komentari kao što su: *pretpostavljeno, isto kao gore, kao i gore, vidi prethodno, vidi XY istraživanje, dodatak, upotrijebljen dijagram toka CONSORT, kineski članak, potreban prijevod, pokus je zaustavljen, nije adekvatno dizajniran, istraživanje je povučeno prije upisa*. Nadalje, sljedeći popratni komentari kategorizirani su isto u ovu kategoriju ako je samo spomenuta potpora za procjenu: *osnovne karakteristike ispitanika; blok randomizacija ili stratifikacija; randomizacija omotnicama, središnja randomizacija po odjelu za statistiku,*

ljekarni ili trećoj strani bez opisa korištene metode. Na temelju ovih pratećih komentara stvorili smo šest potkategorija u kategoriji "metoda nasumičnog odabira nije opisana". Neki od ovih popratnih komentara ukazuju na određene aspekte metodologije povezane s generiranjem nasumičnog redoslijeda, ali nedovoljno dovoljno da bi se RoB mogao na pravilan način procijeniti. Stoga smo stvorili ukupno 19 pod-kategorija popratnih komentara za procjenu, koje su grupirane u sljedećih 5 kategorija: i) tablica nasumičnih brojeva, ii) randomizacija putem računala / softvera / interneta (koja pretpostavlja upotrebu elektroničke automatizacije kao u IVRS-u ili u generatoru nasumičnih brojeva ili pretpostavlja uporabu složenih algoritama kao u minimizaciji), iii) mehanička metoda randomizacije (poput bacanja novčića, izvlačenja lutrije, izvlačenja izmiješanih karata ili omotnica i bacanja kockica), iv) nepotpuna ili neodgovarajuća randomizacija, v) metoda randomizacije nije opisana.

3. Metode rada o prikladnosti procjena Cochraneovih autora za domenu zasljepljivanja ispitanika i osoblja

3.1. Procjena RoB za domenu zasljepljivanja ispitanika i osoblja

Prema Cochraneovom priručniku (odjeljak 8.11.2 i okvir 8.11.a) (89), postavili smo četiri pitanja na koja je trebalo odgovoriti za ispravnu procjenu RoB za domenu zasljepljivanja ispitanika i osoblja : # 1) Tko je bio zaslijepljen?, # 2) Je li zasljepljivanje postignuto?, # 3) Kategorija ishoda?, # 4) Ima li izostanak zasljepljivanja utjecaja na ishod? Koristeći se tim pitanjima i uputama iz Cochraneovog priručnika, a uzimajući u obzir prethodna istraživanja (61, 64, 72, 90-94), prvo smo napravili novu procjenu RoB za domenu zasljepljivanja ispitanika i osoblja, a zatim smo je uspoređivali s procjenama koje su dali autori CSR.

Procjena tko je zaslijepljen

Prema literaturi (61, 72, 90, 91) definirali smo sedam skupina osoba koje bi mogle biti zaslijepljene u istraživanjima: ispitanici, pružatelji zdravstvenih usluga (osoblje), osobe koje prikupljaju podatke, osobe koje bilježe ishode, analitičari podataka (statističari), povjerenstvo za sigurnost podataka i osobe koje pišu tekst istraživanja (63, 64).

Kako bismo analizirali tko je zaslijepljen, usredotočili smo se na dva različita dijela domene pristranosti - ispitanike i osoblje. Pretpostavili smo da su ispitanici bili pravilno zaslijepljeni kad smo pronašli eksplicitan opis zasljepljivanja ispitanika ili opis zasljepljivanja osoblja uz daljnju neizravnu implikaciju zasljepljivanja ispitanika u opisu intervencije, posebno kada se koristio placebo. Ovo "omekšavanje" analize je razumno kada većina napora vezanog za zasljepljivanje ide na zasljepljivanje osoblja (92, 93), s pretpostavkom da je zasljepljivanje ispitanika očito lakše postići. Iz istog razloga, ovo se pravilo nije koristilo *vice versa*, pa je zasljepljivanje osoblja moralo biti jasno navedeno ili nedvojbeno vidljivo iz opisa intervencije.

Nadalje, veliku pažnju posvetili smo spominjanju izraza kao što su: placebo (usto engl. *dummy, sham*) ili „dvostruko-slijepi" (engl. *double-blind, triple-blind, N-blind*). Korištenje izraza *double-blind* isl. kao samostojećih komentara moralo je biti potkrijepljeno detaljnim opisom intervencije kako bi se otkrilo tko je zapravo bio zaslijepljen (64, 72), inače se smatralo da to nije adekvatno navedeno. Isto smo pravilo primijenili i za spominjanje placeba. Iznimno je važno naglasiti da u našoj analizi nismo gledali samo popratni komentar procjeni, već i opis intervencije.

Procjena je li postignuto zasljepljivanje

Smatrali smo da je zasljepljivanje postignuto samo ako su i ispitanici i osoblje bili zaslijepljeni. Kada je opis zasljepljivanja naveden u negativnom tonu (npr. nije urađeno,

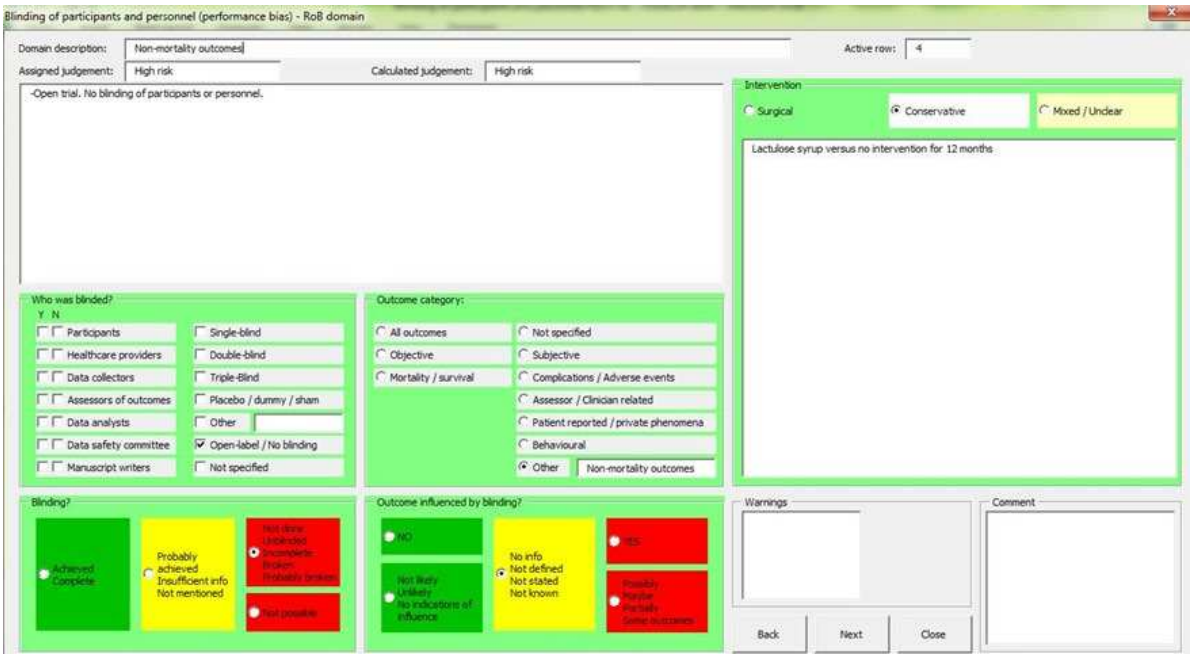
otvoreno istraživanje, nepotpuno zasljepljivanje, prekinuto zasljepljivanje), s vjerojatnošću negativnog tona (npr. vjerojatno nije učinjeno, pretpostavlja se da nije urađeno, vjerojatno prekinuto itd.), ili je samo navedeno djelomično zaslijepljeno, smatrali smo da to nije učinjeno. Svaka druga vrsta opisa smatrana je nejasnom. To je značilo da su se popratni komentari poput „zasljepljivanje vjerojatno postignuto“, „dvostruko slijepi pokus (*double-blind*)“ ili „placebo“, smatrali nejasnima za cijelu domenu. Također smo postavili pravilo da postoji nejasan rizik od pristranosti o zasljepljivanju ispitanika i osoblja ako su autori Cochranea naveli da nema dovoljno podataka ili uopće nisu pružili nikakve podatke.

Procjena kategorije ishoda i je li na ishod utječe izostanak zasljepljivanja

Definicija vrsta ishoda i utjecaj zasljepljivanja na ishode učinjena je izravno. Prema Moustgaard i sur. (94), korišten je izbor od 10 mogućih kategorija ishoda: svi ishodi, objektivni, subjektivni, nisu navedeni, smrtnost/preživljenje, komplikacije/nuspojave, povezani s procjeniteljem/kliničarom, prijavljeni od pacijenta/privatni fenomeni, ponašanje i ostalo. Utjecaj zasljepljivanja na analizirane ishode morao se jasno navesti kao pozitivan ili negativan. Kad se ishodi u tablici navode kao cjelina (svi ishodi), onda je selektivna procjena autora CSR o samo nekim ishodima (npr. navod da nema utjecaja na objektivne ishode) procijenjena kao da će izostanak zasljepljivanja vjerojatno imati utjecaja na sve ishode. Kad utjecaj zasljepljivanja nije bio poznat, niti naveden niti definiran, procijenili smo ga s nejasnim rizikom.

3.2. Razvoj, testiranje i prilagodba raščlanjivača

Konačan izgled MS Excel knjige, nakon ekstrakcije podataka, je bio pregledni radni list koji je integrirao RCT-ove u redcima, a imena domena, procjene rizika, pripadajući komentare i opis intervencije u stupcima za svako izdvojeno istraživanje. Rezultat je tablica sa oko 10500 x 50 polja. S druge strane, procjena rizika od pristranosti o zasljepljivanju ispitanika i osoblja vodi se uputama koje se nalaze u Cochraneovom priručniku. Relevantne informacije za treću domenu vade se iz teksta pratećeg komentara (kako bi se moglo procijeniti je li komentar ispravno potkrijepio procjenu Cochraneovog autora). Jedan komentar može sadržavati nekoliko relevantnih kategorija informacija. Nepreglednost tablice i mnoštvo podataka koji se moraju prikupiti iz komentara zahtijevala je izradu alata koji u jednom vizualnom polju autoru koji analizira podatke omogućuje ciljanu pretvorbu kompleksnog prirodnog govornog jezika u varijable (engl. *parsing*). Na Slika 3 je kao primjer naveden izgled našeg raščlanjivača (parsera) za domenu zasljepljivanju ispitanika i osoblja.



Slika 3 – Izgled korisničkog sučelja raščlanjivača korištenog u našim istraživanjima

Ulazni podatci za parser jesu alfanumeričke vrijednosti u ćelijama MS Excel tablica. Dvostruki klik na bilo koje polje u retku otvara korisničko sučelje parsera. Vrijednosti iz ćelija označenog retka prvo se učitavaju te se potom prikazuju u sučelju parsera, koji osim što ih preglednije prezentira dopušta i njihovu izmjenu. Ove izmjene u sučelju potom mogu na zahtjev korisnika biti upisane kao alfanumeričke vrijednosti u iste ćelije tablice (izlazni podatci).

Za promatranu domenu parser prikazuje dijelove RoB tablice (procjenu rizika i popratni komentar koji potkrepljuje procjenu) te opis intervencije. Pomoću parsera autor na ovaj način ispunjava elektronički obrazac za svaku analiziranu domenu. Dakle, parser samo prikazuje podatke koje autori analiziraju; parser ne pravi nikakve analitičke odluke i nikakve procjene u opisanim istraživanjima ne očekuju se od računalnog programa.

Za treću domenu (zasljepljivanje ispitanika i osoblja) procjena pomoću parsera rađena je na sljedeći način:

- Prvo se ispunjava dio koji ima unaprijed zadane vrijednost o tome tko je prema komentaru bio zaslijepljen (ispitanik, pružatelj intervencije, osoba koja prikuplja podatke o ishodima, analitičari podataka, autori teksta članka itd.). Potom, koristeći upute iz Cochraneovog priručnika odgovara na ključno pitanje je li zasljepljivanje postignuto (potvrdno, vjerojatno/moguće ili upitno/nepostignuto/nemoguće).
- Sljedeći korak jest kategorizacija ishoda u unaprijed zadane kategorije (subjektivni, objektivni, vezani za kliničara, vezani za ispitanika, svi ishodi).

- Naposljetku odgovara na pitanje ima li izostanak zasljepljivanja utjecaj na ishod (ima, upitno, nema).
- Zaključno, prema opisu intervencije autor procjenjuje radi li se o kirurškoj/invazivnoj, farmakološkoj/psihosocijalnoj ili miješanoj intervenciji.

Pristupnik je testirao parser na uzorku od 500 RCT-ova. Testiranje parsera je opet bilo na razini alfa testiranja kako bi se otklonile greške samog parsera. I u ovom slučaju u konsenzusu s mentoricom proširene su mogućnosti upisa podataka kroz sam alat.

U istraživanju o domeni o zasljepljivanju ispitanika i osoblja mentorica je provjerila pilot kategorizaciju od 500 procjena koje je uradio doktorand. Doktorand je za provjeru mentorici ostavio 33 svoja komentara vezanih za vlastite procjene kao i dodatnih 12 komentara vezanih za postizanje zasljepljivanja. Komentari doktoranda su u ovoj ranoj fazi istraživanja otkrili zbunjujuće faktore poput korištenja izraza *double-blind* i/ili placebo u formi samostojećeg ili prekratkog komentara (s malo znakova). Mentorica je tada predložila pravila vezana za postavljene komentare koje smo dosljedno pratili i prenijeli ko-autorima. Zbog utvrđenih pravila urađeno je više istovjetnih izmjena što je rezultiralo s ukupno 77 promjena u 500 procjena.

Upravo nakon što su otklonjene osnovne greške te je kroz sustav pristupnik uspio obraditi cijeli randomizirani uzorak od 500 istraživanja uradili smo i, u razvojnom smislu, beta-testiranje na način da je omogućeno korištenje parser alata malom broju korisnika (jednom koautoru) u pravom okruženju – s pravim podatcima, a zauzvrat smo dobili komentare o poteškoćama koje su postojale. Na temelju beta-testiranja održan je zajednički sastanak gdje je prikazan rad u sučelju te obrađena ključna pitanja proizišla iz alfa i beta testiranja.

Po obradi ostalih RCT-a od strane ostalih koautora, tekst je potpuno preveden u varijable te je korištenjem jednostavne formule u MS Excel tablici *(„=IF(OR(Z5>0,01;AA5>0,01);"Low risk";IF(Z5*AA5=0,0001;"Unclear risk";"High risk"))“*) dobivena točna procjena rizika za ukupno 7151 RCT. Ona je temeljena samo na pitanjima o postignutom zasljepljivanju i utjecaju na ishod, a prema tablici 8.5.d Cochraneovog priručnika.

Ovom formulom se u MS Excel polju materijalizira jednostavno zaključivanje na temelju dviju brojčanih vrijednosti (premise zaključka). Premise su vidljive u tablici *Blinding of outcome assessment - final splitting outcome.xlsm* i to u stupcima Z# (čiji je naslov *Blinding achieved*) i AA# (čiji je naslov *Outcome influenced*). U stupcu Z koristi se skup brojeva umjesto tekstualnih varijabli (-2; -1; 0,01; 1) koji odgovara odgovorima procjenitelja kroz parser (Upute za parser.pptx sličica broj 5) i to redom (*blinding not possible; blinding not*

done/unblinded/incomplete/broken; blinding probably achieved/insufficient info/not mentioned; achieved/complete). Za stupac AA o utjecaju na ishode koriste se brojčane varijable (-2; -1; 0,01; 1; 2) i to odgovara tekstu kao na sličici 7 u uputama parsera. Korištenje brojčanih varijabli dopušta prečace u programskom kodu i zbog toga su korištene jer je formula jednostavnija. Pitanje zbog čega u slučaju ne-negativnosti ili ne-pozitivnosti koristimo broj 0,01 umjesto 0 jest pitanje umnoška. I negativan i pozitivan broj pomnožen s nulom daju nulu (ne-pozitivan i ne-negativan broj) stoga raščlamba formiranjem umnoška ova dva koeficijenta (Z i AA) daje 0,0001 samo ako su oba člana 0,01 i tada je ocjena rizika „*unclear*", a ne i svaki put kad bi jedna varijabla bila jednaka nuli onda bi i zaključak bio „*unclear risk*" za svaki takav slučaj. Stoga pri korištenju broja tek neznatno većeg od nule dobivamo negativan broj koji nam označava visoki rizik, a pozitivan broj veći od 0,0001 koji označava nizak rizik. Osim u predmetnim tablicama na OSF objašnjenje je postavljeno i u obliku komentara programskog koda u datoteci s kodom za *UserForm* parsera na OSF: *06 - Parser - UserFormUI.bas*.

Samo sučelje parsera je ustrojeno na način da ne dopušta odstupanja od definiranog okvira pa tako ni preskakanje važnih elemenata na koje zaboravljaju i autori sustavnih pregleda (prema rezultatima naših istraživanja). Isto tako, parser korisniku postavlja četiri pitanja, a logički zaključak se izvodi iz ključna dva. Ovaj višak informacija (redundancija) koji se odnosi raščlambu subjekata i ishoda iskorišten je za otklanjanje pogreški autora prilikom raščlambe kao i utvrđivanje uzroka nepodudarnosti procjena rizika od pristranosti.

Dodatno korištenje redundancije je bilo prilikom utvrđivanje vrste intervencije na više razina: koautori su odabirali vrstu intervencije za svaki RCT prilikom raščlambe, a u randomiziranom uzorku velik broj RCT-a se ponavljao i dobio jednako toliko procjena od različitih autora. S druge strane u analizi iduće domene svaki je od koautora dobio novi dio uzorka istih RCT-ova i ponovno im odredio vrstu intervencije tako da su bez spoznaje o tome koautori procjenu intervencija cijelog uzorka odradili dvaput.

U istraživanju o domeni o zasljepljivanju ispitanika i osoblja ko-autori su uradili ponovnu procjenu rizika pristranosti za 6918 popratnih komentara. Za dodatnu analizu ostavili su ukupno najmanje 161 komentar na pregled doktorandu. Razlika početnih procjena ko-autora i kasnije provjere doktoranda bila je u 706 promjena u pitanju je li postignuto zasljepljivanje, dok je urađeno 80 promjena vezanih za pitanje osjetljivosti ishoda na izostanak zasljepljivanja. To je rezultiralo s ukupno 661 (9,6%) promjena u kategoriji rizika pristranosti. U usporedbi s neprovjerenim procjenama ko-autora finalna nepodudarnost koja je navedena u rezultatima istraživanja bila je 1828 od 6918 (26,4%), a da korekcije nije bilo iznosila bi 1933 (27,9%).

Napominjemo da je u opisanim istraživanjima računalni program korišten isključivo kao alat koji značajno olakšava a) ekstrakciju podataka, b) tablični unos procjena procjenitelja prema svim definiranim informacijama. Računalni program za ekstrakciju podataka ni u kom slučaju ne utječe na procjenu procjenitelja upravo jer je korišten isključivo za olakšavanje ponavljanih, vremenski zahtjevnih i zamarajućih repetitivnih radnji prilikom ekstrakcije: pregledavanje, klizanje, označavanje, kopiranje i lijepljenje (engl. *viewing, scrolling, selecting, copying, pasting*). Računalno sučelje koje je osmišljeno za bilježenje procjena procjenitelja isto tako nije moglo utjecati na procjenu procjenitelja jer im ne nudi nikakva rješenja, nego im prezentira informacije koje se nalaze u komentaru iz RoB tablica, i traži od procjenitelja da odgovore na pitanja koja se tiču procjene tih domena. K tome sučelje ne dopušta nepotpuni unos podataka, a postojanje nepravilnosti/nelogičnosti lako se uoči jer procjenitelj mora obraditi svaki pojedini slučaj od početka do kraja. Stoga se u računalnim rješenjima koja su korištena u istraživanju ne nudi procjeniteljima ništa što bi na njih utjecalo ili umjesto njih odlučivalo. Još jednom naglašavamo da je raščlamba urađena ručno za sve procjene u svim radovima, a da je to za radove o zasljepljivanju urađeno kroz *UserForm* parsera u MS Excelu. Sva je daljnja obrada urađena ručno u MS Excel knjizi.

Kako bismo i ove tvrdnje učinili transparentnima kodove makronaredbi i izrađene dijagrame toka makronaredbi uz popratne komentare o logičkim pravilima postavili smo na OSF https://osf.io/fmjxz/.

3.3. Primarni ishodi

Analizirali smo je li broj i podudarnost procjena RoB za domenu zasljepljivanja ispitanika i osoblja koju su dodijelili autori Cochrane u skladu s preporukama iz Cochrane priručnika. Priručnik je korišten kao zlatni standard u našoj procjeni; ako se procjene autora CSR nisu u potpunosti pridržavale smjernica iz Cochraneovog priručnika, smatrali smo ih nepodudarnima.

3.4. Sekundarni ishodi

Analizirali smo krajnji rezultat korištenja nedefiniranih samostojećih izraza *double-blind* i placebo (nesigurnost u potkrepljujućim komentarima), utjecaj duljine pratećeg komentara (engl. *length of comment - LOC*) za procjenu dodijeljenu od strane autora CSR u odnosu na ponovnu procjenu od strane naše grupe, utjecaj razdvajanja procjene prema različitim ishodima (podjela ishoda) i utjecaj dijeljenja zasljepljujućih subjekata na pod-domene (dijeljenje domene).

3.5. Statistički postupci

Opisni podaci predstavljeni su kao frekvencije i postotci. Pogreška tipa I $\alpha = 0,05$ i pogreška tipa II $\beta = 0,2$ korištene su u svim statističkim postupcima. Normalnost skupova podataka ispitana je primjenom Kolmogorov-Smirnovljeva testa. Za usporedbu neovisnih skupina neparametrijskih podataka korišten je Mann-Whitneyev test; za više od dvije skupine podataka korišten je Kruskal-Wallisov test. Za ovisne skupine neparametrijskih podataka korišten je Wilcoxonov test (upareni uzorci). Statističke analize provedene su pomoću računalnog programa *MedCalc za Windows*, verzija 12.5.0.0 (*MedCalc Software*, *Ostend*, Belgija). Detaljni opisi korištenih testova nalaze se u Tablica 23, u odjeljku dodatni materijali.

4. Metode rada o prikladnosti procjena Cochraneovih autora za domenu rizika pristranosti o zasljepljivanju osoba koje bilježe ishode

Za četvrtu RoB domenu o zasljepljivanju osoba koje bilježe ishode koristili smo istovjetan parser kao i za treću domenu. Testnu (beta) analizu uradila je druga autorica na uzorku od 500 RCT-a, a analizu je provjerio pristupnik. Obradu ostalih RCT-a uradili su koautori te je analizirano ukupno 7626 RCT-a.

S obzirom da je pilot kategorizaciju radila druga autorica, a kontrolu doktorand kvantifikacija pogreške je opisana sumarno. U istraživanju o domeni o zasljepljivanju osoba koje bilježe ishode ko-autori su uradili ponovnu procjenu rizika pristranosti za 8656 popratnih komentara. Za dodatnu analizu ostavili su najmanje 857 komentara na pregled doktorandu. Razlika početnih procjena koautora i kasnije provjere doktoranda bila je u 368 promjena u pitanju je li postignuto zasljepljivanje, dok je urađeno 974 promjena vezanih za pitanje osjetljivosti ishoda na izostanak zasljepljivanja. To je rezultiralo s ukupno 894 (10,3%) promjene u kategoriji rizika pristranosti. U usporedbi s neprovjerenim procjenama ko-autora finalna nepodudarnost koja je navedena u rezultatima istraživanja bila je 1908 od 8656 (22,0%), a da korekcije nije bilo iznosila bi 2378 (27,5%).

4.1. Procjena RoB za domenu zasljepljivanja osoba koje bilježe ishode

U istovjetnom korisničkom sučelju napravili smo novu procjenu RoB za RCT u kojoj su autori CSR naveli sve potrebne informacije u RoB tablici za domenu o zasljepljivanju osoba koje bilježe ishode, tj. i procjenu (nizak, visok ili nejasan RoB) i popratni komentar. Slijedili smo upute za procjenu iz Cochraneovog priručnika (odjeljak 8.12.2.) (89) i definirali da je potrebno pravilno odgovoriti na dva glavna pitanja kako bi se adekvatno procijenio RoB. Pitanje # 1 bilo je: tko procjenjuje ishod? - jer se mogu koristiti različite osobe koje bilježe ishode i mora se procijeniti jesu li bili zaslijepljeni ili ne. Pitanje br. 2 glasilo je: postoji li rizik od pristranosti zbog tipa ishoda? - jer nisu svi ishodi pod jednakim utjecajem izostanka zasljepljivanja osoba koje bilježe te ishode. Na primjer, za ishod poput ukupne smrtnosti, nedostatak zasljepljivanja osoba koje bilježe ishod neće utjecati na ishod. Kako bi procijenili jesu li autori CSR spomenuli vrstu ishoda koja je procijenjena unutar analizirane domene o zasljepljivanju osoba koje bilježe ishode, kategorizirali smo svaki ishod u jednu od šest unaprijed definiranih kategorija: i) objektivno mjereni / nesubjektivni, ii) ishodi vezani uz kliničara, iii) ishodi povezani s pacijentom (samoprijavljeni), iv) subjektivni ishodi, v) svi ishodi, vi) ishodi nisu navedeni, na temelju razmatranja koliko je ishod bio subjektivan ili

objektivan (94). Posljednje dvije kategorije ne odnose se na određenu vrstu ishoda na isti način kao i prve četiri. 'Svi ishodi' znače da su autori zajedno prosuđivali rizik pristranosti za sve ishode, tj. nisu naveli određene podskupine ishoda za ovu RoB domenu. Kategorizirali smo da ishodi 'nisu navedeni' kad autori u RoB tablici nisu spomenuli nijedan određeni ishod tipa s imenom domene. Nenavođenje ishoda koji domena procjenjuje zadana je postavka *RevMan* (95), softvera koji se koristi za izradu CSR. Na kraju, usporedili smo naše nove procjene s procjenama autora CSR.

4.2. Primarni ishodi

Procjene za RoB u domeni o zasljepljivanju osoba koje bilježe ishode koje su dodijelili autori CSR analizirane su prema broju, vrsti i podudarnosti. Standard definicije u našoj, ponovljenoj, procjeni bio je Cochraneov priručnik. Smatrali smo da je procjena autora CSR nepodudarna ako se nije u potpunosti pridržavala smjernica Cochraneovog priručnika tj. nije bila podudarna s našom.

4.3. Sekundarni ishodi

Analizirali smo pojavnost razdvajanja RoB domene o zasljepljivanju osoba koje bilježe ishode prema vrsti ishoda(jesu li Cochraneovi autori ovu RoB domenu razdvojili na dvije ili više pod-domena), vrste ishoda prema kojima je korišteno razdvajanje i adekvatnost (podudarnost) procjena RoB u različitim kategorijama ishoda.

4.4. Statistički postupci

Sve smo opisne podatke prikazali kao frekvencije i postotke. Za sva statistička ispitivanja koristili smo pogrešku tipa I = 0,05 i pogrešku tipa II $\beta = 0,2$. Statističke analize provedene su pomoću računalnog programa *MedCalc za Windows*, verzija 12.5.0.0 (*MedCalc Software*, Ostend, Belgija). Svi su skupovi podataka testirani na normalnost Kolmogorov-Smirnovljevim testom. Mann-Whitneyev test korišten je za usporedbu neovisnih uzoraka neparametrijskih podataka, a Wilcoxon test za uparene uzorke. Razlika u omjerima ispitana je hi-kvadrat testom. Hipoteze, mjere ishoda, korišteni statistički testovi i njihovi rezultati dostupni su u Tablica 24, u odjeljku dodatni materijali.

5. Metode rada o istraživanju o kirurškim intervencijama

5.1. Koncept istraživanja i protokol

Ovo je bila primarno metodološko istraživanje, koje je analiziralo metode korištene u objavljenim Cochraneovim sustavnim pregledima u skladu sa *STREngthening the Reporting of Observational Studies in Epidemiology* (96) (STROBE Statement).

5.2. Kriteriji za uključivanje i isključivanje

Analizirani su Cochraneovi pregledi objavljene između srpnja 2015. i lipnja 2016. u Cochraneovoj bazi podataka sustavnih pregleda (*CDSR*). Ovo je bio prikladan jednogodišnji uzorak zasnovan na našim prethodnim istraživanjima (82-84, 87), iz razdoblja od četiri godine nakon uvođenja RoB alata 2011. godine. CSR koji su uključivali samo RCT-ove ili i RCT-ove i nerandomizirana istraživanja bile su prihvatljive. Isključeni su dijagnostički, prazni ili povučeni pregledi, kao i oni koji su uključivali samo nerandomizirana istraživanja.

5.3. Definicija i kategorizacija intervencija

Pokusi su kategorizirani kao kirurški, konzervativni ili mješoviti, ovisno o invazivnosti njihove intervencije (ili komparatora). U svrhu ovog istraživanja, invazivnost se smatrala nečim što zahtijeva prisan kontakt između osobe koja obavlja postupak i osobe koja treba postupak. Strogo invazivni (ili kirurški) postupci su medicinski postupci koji ulaze u tijelo, obično rezanjem ili probijanjem organa (prvenstveno kože, ali ostalih organa, kao i sluznice, zuba itd.) Ili umetanjem instrumenata u tijelo kroz ove rezove. Osim otvorenih kirurških zahvata, ova kategorija uključuje usporedbe različitih kirurških tehnika, stomatoloških intervencija, ERCP (endoskopska retrogradna kolangiopankreatografija) ili EMS (ekstramukozna resekcija) tehnika.

Manje invazivni (nejasni ili mješoviti) postupci uključuju ulazak u tjelesnu šupljinu ili prekid normalnih tjelesnih funkcija i mogu uključivati probijanje kože, davanje ne-oralnih lijekova, umetanje cjevčice ili medicinskih uređaja i njegu nakon medicinskih postupaka, poput njege stome i njege katetera. Primjeri su akupunktura, vanjska manipulacija fetusom (ECV - vanjska cefalična rotacija), različiti modaliteti umjetne ventilacije / disanja (npr. CPAP - kontinuirani pozitivni tlak u dišnim putovima ili HFJV - visokofrekventna mlazna ventilacija), vrste anestezije (blokade, spinalna, opća), primjena različitih vrsta katetera bez promjene u tehnici primjene (osim ako je tehnika jasno kirurška), intervencije koje simuliraju invazivne tehnike

(virtualni trener), invazivni postupak koji nije strogo kirurški (punkcije, intraartikularna injekcija).

Svaka druga intervencija bez dokaza o invazivnosti smatrana je konzervativnom ili neinvazivnom. To također uključuje intervencije ili različite postupke prije ili nakon kirurškog liječenja (različita fizikalna terapija ili lijekovi) koji ne mijenjaju izvedbu kirurškog postupka koji se planira ili primjenjuje u tom trenutku.

Intervencije su kategorizirala dva autora. Prva kategorizacija intervencija izvršena je tijekom analize domene za zasljepljivanje ispitanika i osoblja; radi provjere, kategorizacija je ponovljena neovisno tijekom analize domene za zasljepljivanje osoba koje bilježe ishode. Neskladnosti i nedostajuće kategorizacije razriješio je prvi autor.

Odluku o spajanju mješovitih i ne-kirurških kategorija donijeli smo zbog uočenog slaganja dviju neovisnih kategorizacija i najboljeg konsenzusa (suglasja, engl. *inter-rater agreement*). Naposlijetku, istraživanja su podijeljena prema intervencijama na kirurške ili ne-kirurške. Pojedinosti o kategorizacijama i konsenzusu predstavljene su u Tablica 19 – Konsenzus (suglasje) dvaju procjenitelja i varijabilnost različitih kategorizacija intervencija.

5.4. Procjena primjerenosti za četiri domene alata za rizik od pristranosti

U svakom uključenom RCT iz uključenih CSR urađena je analiza jesu li procjene autora CSR bile podudarne za sljedeće četiri RoB domene: generiranje nasumičnog redoslijeda, prikrivanje razvrstavanja, zasljepljivanje ispitanika i osoblja i zasljepljivanje ocjenjivača osoba koje bilježe ishode. Podudarnost procjena analizirana je uspoređivanjem procjena izvornih Cochrane autora s našim ponovno procijenjenim RoB. Izvor naših procjena bio je popratni komentar iz RoB tablice i opis intervencije koju su dali autori CSR. Za RoB domene za generiranje nasumičnog redoslijeda i prikrivanje razvrstavanja popratni komentari kategorizirani su da se donese procjena kako je opisano u našim prethodnim istraživanjima (83, 84). Za dvije promatrane domene o zasljepljivanju trebalo je utvrditi koji je subjekt zaslijepljen i jesu li ishodi pod utjecajem nedostatka zasljepljivanja (82, 87). U završnoj fazi uspoređena je pojavnost nepodudarnih procjena i navedeni razlozi za nepodudarne procjene RoB-a između istraživanja s kirurškim i ne-kirurškim intervencijama.

5.5. Primarni ishod

Primarni ishod bila je prevalencija nepodudarnih procjena RoB-a za četiri Cochrane RoB domene u istraživanja s kirurškim i ne-kirurškim intervencijama.

5.6. Sekundarni ishodi

Sekundarni ishodi bili su raspodjela procjena RoB (niska / nejasna / visoka) i prevalencija različitih razloga za nepodudarne procjene u istraživanja s kirurškim i ne-kirurškim intervencijama.

5.7. Statistički postupci

Opisni podaci predstavljeni su kao frekvencije i postotci. Prije analize, skupovi podataka testirani su na normalnost raspodjele Kolmogorov-Smirnovljevim testom. Za neparametrijske podatke korišten je Wilcoxonov test za uparene uzorke, Mann-Whitney test za usporedbu dva neovisna uzorka, dok je Kruskal-Wallis test korišten za usporedbu tri ili više uzoraka. Kad je Kruskal-Wallisov test bio pozitivan ($P < 0,05$), provedena je usporedba podskupina prema Conoveru. Nisu uzete u obzir prilagodbe p-vrijednosti u *post-hoc* analizama, s obzirom na to da je naša studija istraživačka i uključuje *post-hoc* testiranje neplaniranih usporedbi (97). Iz istog razloga, s idejom da se naglasi hipoteza u nastajanju koja se može razmatrati za izradu daljnjih istraživanja, ako *post-hoc* analiza nije otkrila razlike u usporedbi, jednosmjerna ANOVA ponovno je primijenjena na skupu podataka umjesto Kruskal-Wallisova testa, dok je za usporedbu podskupina korišten Student-Newman-Keulsov test. Razlika u omjerima ispitana je hi-kvadrat testom. Za sva statistička ispitivanja koristili smo pogrešku tipa I = 0,05 i pogrešku tipa II $\beta = 0,2$. Statističke analize provedene su pomoću *MedCalc* za Windows, verzija 12.5.0.0 (*MedCalc Software*, Ostend, Belgija). Izračunali smo direktnu proporciju konsenzusa i prikazali je zajedno s Cohenovom neponderiranom kappom s odgovarajućim 95% CI (interval pouzdanosti) kao mjeru konsenzusa među autorima (98). Razinu slaganja klasificirali smo na sljedeći način: vrijednosti ≤ 0 kao da ne pokazuju nikakvo slaganje, 0,01 - 0,20 kao nikakvo do malo, 0,21 - 0,40 kao dovoljno, 0,41 - 0,60 kao umjereno, 0,61 - 0,80 kao značajno i 0,81 - 1,00 kao gotovo savršeno slaganje. Ishodi, hipoteze, statistički testovi s odgovarajućim rezultatima i zaključci predstavljeni su u Tablica 25.

REZULTATI

1. Rezultati prvog rada – prikladnost procjena Cochraneovih autora za domenu rizika pristranosti o generiranju nasumičnog redoslijeda

Analizirali smo 10290 RCT-a uključenih u 718 CSR-a. Tijekom analize morali smo isključiti 187 RCT-a i 14 CSR-a (Slika 1). Isključili smo 19 RCT-a zbog različitog ocjenjivanja rizika - prilagođenog za više od tri uobičajene procjene RoB (niska, visoka i nejasna). Dodatnih 112 istraživanja isključena su zbog nedostatka RoB tablice ili domene povezane s generiranjem nasumičnog redoslijeda. Za 49 RCT-a nismo pronašli specifični razlog za nedostatak RoB tablice ili domene, dok za 54 istraživanja nije utvrđeno da su randomizirana, u tri slučaja pogreške na mrežnoj stranici uzrokovale su dvostruke unose, a šest istraživanja imalo je samo sažetke (Tablica 1). Također smo isključili 56 RCT-a s praznim poljem (ili oznakom N/A koja predstavlja "nije dostupno") kao popratnog komentara u RoB tablici.

Tablica 6 - Istraživanja kojima je nedostajala RoB domena o generiranju nasumičnog redoslijeda

Razlog isključenja (ustroj istraživanja)	**N**
Nestandardna procjena RoB	**19**
Domena nedostaje u RoB tablici	**96**
Nerandomizirano istraživanje	50
Randomizirani pokus s paralelnim skupinama	10
Randomizirani kontrolirani pokus	36
RoB tablica nedostaje	**16**
Nerandomizirano istraživanje	4
Randomizirani kontrolirani pokus	3
Dvostruki unos	3
Ustroj istraživanja nejasan – dostupan samo sažetak	6
Ukupno	**131**

1.1. Broj i vrsta različitih pratećih komentara

U našoj glavnoj analizi popratne komentare za procjenu u preostalih 10103 RCT-a svrstali smo u pet kategorija. Korištenjem kriterija iz Cochraneova priručnika, u više od polovice tih RCT-a metoda randomizacije nije opisana (N = 5706), dok su preostali popratni komentari naznačili da je randomizacija izvršena pomoću računala / softvera / interneta (N = 2850), tablica nasumičnih brojeva (N = 883), mehanička metoda randomizacije (N = 359) i nepotpuna ili

neprikladna randomizacija (N = 305). Učestalost ovih kategorija i različite vrste popratnih komentara, koje smo pronašli u svakoj od ovih pet kategorija, popisani su u Tablica 7.

Tablica 7 – Kategorije popratnih komentara za procjenu RoB u domeni generiranja nasumičnog redoslijeda

Categorije popratnih komentara za procjenu rizika od pristranosti povezanu s generiranjem nasumičnog redoslijeda	**Ukupno, N (%)**	**Visok rizik, N (%)**	**Nejasan rizik, N (%)**	**Nizak rizik, N (%)**
Tablica nasumičnih brojeva (nizak rizik†)	**883 (8,7)**	**11 (1,2)**	**31 (3,5)**	**841 (95,2)**
tablica nasumičnih brojeva	883 (100,0)	11 (1,2)	31 (3,5)	841 (95,2)
Randomizacija pomoću računala/računalnog programa/interneta (nizak rizik†)	**2850 (28,2)**	**2 (0,1)**	**36 (1,3)**	**2812 (98,7)**
računalo/računalni program	2459 (86,3)	0 (0,0)	17 (0,7)	2442 (99,3)
IVRS*	60 (2,1)	0 (0,0)	7 (11,7)	53 (88,3)
minimizacija	103 (3,6)	2 (1,9)	8 (7,8)	93 (90,3)
generator nasumičnih brojeva (bez spomena računala)	98 (3,4)	0 (0,0)	0 (0,0)	98 (100,0)
internet	130 (4,6)	0 (0,0)	4 (3,1)	126 (96,9)
Mehanička metoda randomizacije (nizak rizik†)	**359 (3,6)**	**7 (1,9)**	**37 (10,3)**	**315 (87,7)**
bacanje novčića	85 (23,7)	0 (0,0)	1 (1,2)	84 (98,8)
izvlačenje brojeva	201 (56,0)	4 (2,0)	12 (6,0)	185 (92,0)
miješanje kartica ili kuverti	60 (16,7)	3 (5,0)	24 (40,0)	33 (55,0)
bacanje kocke	13 (3,6)	0 (0,0)	0 (0,0)	13 (100,0)
Nepotpuna ili neprikladna randomizacija (visok rizik†)	**305 (3,0)**	**264 (86,6)**	**32 (10,5)**	**9 (3,0)**
kvazi-randomizacija	238 (78,0)	209 (87,8)	22 (9,2)	7 (2,9)
nepotpuna, pogrešna randomizacija	41 (13,4)	34 (82,9)	7 (17,1)	0 (0,0)
procjena, preferencija, dostupnost	26 (8,5)	21 (80,8)	3 (11,5)	2 (7,7)
Metoda randomizacije nije opisana (nejasan rizik†)	**5706 (56,5)**	**81 (1,4)**	**4651 (81,5)**	**974 (17,1)**
nije opisano	4586 (80,4)	69 (1,5)	4272 (93,2)	245 (5,3)
blok-randomizacija/stratifikacija	679 (11,9)	2 (0,3)	219 (32,3)	458 (67,5)
kuverte	247 (4,3)	3 (1,2)	117 (47,4)	127 (51,4)
središnja randomizacija (statistički odjel, ljekarna, treća strana)	124 (2,2)	1 (0,8)	0 (0,0)	123 (99,2)
neravnoteža u početnim osobinama između skupina	41 (0,7)	6 (14,6)	35 (85,4)	0 (0,0)
ravnoteža u početnim osobinama između skupina	29 (0,5)	0 (0,0)	8 (27,6)	21 (72,4)
Ukupno, N (%)	**10103 (100,0)**	**365 (3,6)**	**4787 (47,4)**	**4951 (49,0)**

**IVRS = Interactive Voice Response System (interaktivni sustav glasovnog odgovora). †Kategorija RoB procjene prema Cochraneovom priručniku.*

1.2. Podudarnost procjene RoB povezan s generiranjem nasumičnog redoslijeda

Većina od pet kategorija popratnih komentara imala je točne procjene Cochraneovih autora. U kategoriji u kojoj metoda randomizacije nije opisana, 82% istraživanja ispravno je ocijenjeno kao nejasno. Za tablicu nasumičnih brojeva, 95% Cochraneova autora ispravno je to procijenilo kao nizak RoB. Kada su Cochraneovi autori naznačili da postoji nepotpuna ili neprimjerena randomizacija, 87% takvih istraživanja ispravno je ocijenjeno kako imaju visok RoB povezanog s generiranjem nasumičnog redoslijeda. Za popratne komentare koji opisuju randomizaciju putem računala / računalnog programa / interneta, 99% autora ispravno je ocijenilo da je to nizak RoB. Za mehaničke metode randomizacije, 88% autora pravilno je ocijenilo da se radi o istraživanjima koja imaju nizak RoB u ovoj domeni.

Ukupno gledano, 1220/10103 istraživanja (12%) pogrešno su procijenjena zbog rizika od pristranosti povezanog s generiranjem nasumičnog redoslijeda. Najveći udio pogrešaka uočen je u kategoriji „metoda randomizacije nije opisana“. Na primjer, 99% istraživanja u kojima je popratni komentar samo ukazivao na središnju randomizaciju pogrešno je ocijenjeno kao nizak rizik od pristranosti jer metoda randomizacije zapravo nije opisana. Isto tako, 72% pokusa koji su pisali samo o početnoj ravnoteži između skupina, 68% pokusa koji su imali samo opis randomizacije u blokovima ili stratifikacije te 51% pokusa za koja je poprani komentar spomenuo samo da je randomizacija urađena pomoću omotnica, ocijenjeno je niskim RoB (Tablica 7).

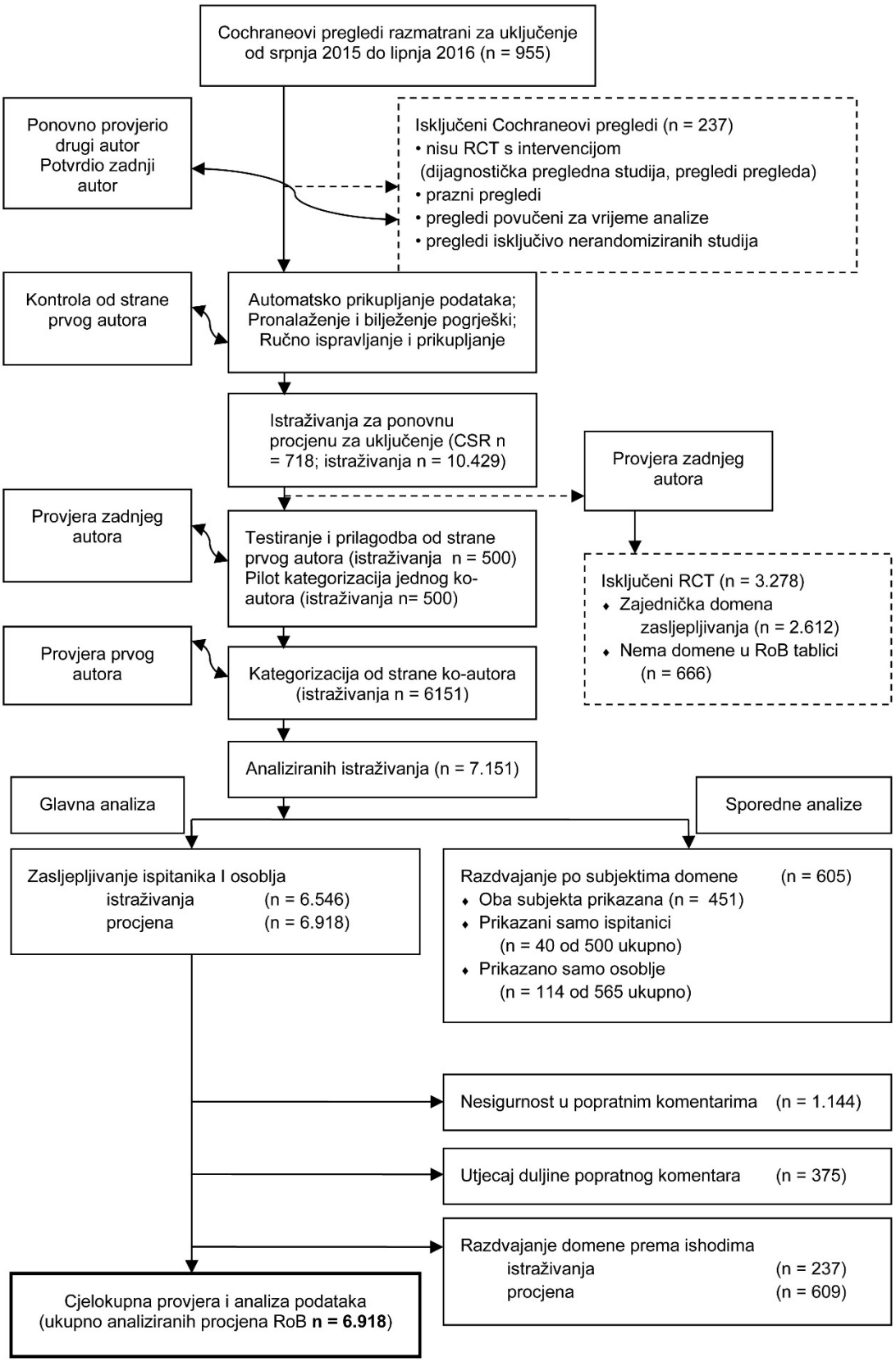
Najveći udio pogrešaka u procjenama RoB utvrđen je u RCT-ima koji su kategorizirani kao da imaju visok rizik od pristranosti povezanih s generiranjem nasumičnog redoslijeda. Bilo je 365 (3,6%) RCT-a za koje je ocijenjeno da imaju visok rizik od pristranosti za ovu domenu, od čega je 101 (28%) pogrešno ocijenjeno. Od pet kategorija koje smo koristili, samo kategoriju „nepotpune ili neodgovarajuće randomizacije″ trebalo je ocijeniti visokim rizikom od pristranosti (Tablica 7).

Među RCT-ima za koje se procjenjuje da imaju mali rizik od pristranosti za generiranje nasumičnog redoslijeda (N = 4951, 49%), bilo je 983 (20%) s pogrešnom procjenom. Među pet kategorija, mali rizik od pristranosti trebao bi biti povezan s upotrebom tablice nasumičnih brojeva, randomizacijom putem računala / softvera / interneta i mehaničkim metodama randomizacije.

Za 4787 RCT-a (48%) procijenjeno je da imaju nejasan RoB za domenu generiranja nasumičnog redoslijeda, od čega je 136 (3%) pogrešno ocijenjeno. Od pet nadređenih kategorija koje smo koristili, samo komentari u kategoriji gdje metoda randomizacije nije opisana trebali bi biti procijenjeni da imaju nejasan RoB.

2. Rezultati drugog rada – prikladnost procjena Cochraneovih autora za domenu zasljepljivanja ispitanika i osoblja

Analizirali smo RoB tablice od 10429 RCT-a uključenih u 718 CSR-ova (Slika 4). Domenu o zasljepljivanja ispitanika i osoblja, kako je definirana u Cochraneovom priručniku, pronašli smo i analizirali u 6546 RCT-a. Usto smo pronašli 237 RCT-a u kojima su napravljene procjene za više od jednog ishoda (dijeljenje prema ishodima), što je rezultiralo s ukupno 6918 procjena.



Slika 4 – Dijagram tijeka uključivanja i isključivanja istraživanja

U sekundarnim analizama pronašli smo dodatnih 605 RCT-a u kojima su autori podijelili domenu na zasljepljivanje ispitanika (n = 500) ili zasljepljivanje osoblja (n = 565). Većina njih (n = 451) obuhvaćala je oba dijela podijeljene domene.

2.1. Broj i podudarnost procjena RoB za domenu zasljepljivanja ispitanika i osoblja

U našoj glavnoj analizi razmatrano je ukupno 6918 procjena RoB-a iz Cochraneovih pregleda. Najveći udio procjena, 2574 (37,2%), bio je visokog rizika, zatim 2333 (33,7%) niskog rizika, a 2011 (29,1%) nejasnog RoB (Tablica 8).

Tablica 8 – Podudarnost procjena RoB za domenu zasljepljivanja ispitanika i osoblja

Procjena objavljena u Cochraneovim pregledima	**Nova procjena izračunata po Cochraneovom priručniku**			**Ukupno N (%)**
	Visok rizik	**Nizak rizik**	**Nejasan rizik**	
Visok rizik	2386	5	183	**2574 (37,2%)**
Nizak rizik	96	1094	1143	**2333 (33,7%)**
Nejasan rizik	375	26	1610	**2011 (29,1%)**
Ukupno	**2857 (41,3%)**	**1125 (16,3%)**	**2936 (42,4%)**	**6918 (100,0%)**
Neprikladne procjene	**471 (16,5%)**	**31 (2,8%)**	**1326 (45,2%)**	**1828 (26,4%)**

Naš izračun RoB-a za domenu zasljepljivanja ispitanika i osoblja pokazao je značajnu razliku (p <0,0001, Tablica 23, dodatni materijali) u između originalnih procjena RoB od strane autora CSR i naših ponovljenih procjena. Otkrili smo da su autori CSR nepodudarno procjenjivali RoB u 1828 slučajeva (26,4%). Naša je analiza dala „lošiju“ procjenu RoB u 1614 (88,3%) tih slučajeva (promjena procjene RoB iz niske u nejasnu ili iz nejasne u visoku) i „bolju“ RoB procjenu u 214 (11,7%) slučajeva (promjena procjene iz nejasne u nisku ili iz visoke u nejasnu), kao što je prikazano u Tablica 9. Otkrili smo da je 7,3% procjena visokog RoB koje su dali autori CSR ponovno procijenjeno kao nejasan ili nizak RoB, dok je 19,9% dodijeljenih nejasnih procjena RoB bilo izračunato ili kao visok ili kao nizak RoB. Polovica procjena (53,1%) kojima je dodijeljen nizak RoB bila je ponovno procijenjena kao nejasan ili visok RoB.

Tablica 9 – Razlike u procjenama koje su dali autori CSR i procjena u skladu s Cochraneovim priručnikom (ukupno 6918 procjena)

Promjena smjera procjene RoB	**N (%)**		**Medijan duljine komentara (znakovi)**	**95% CI**
Prema višem RoB	**1614**	**(23,3%)**		
Nizak u Visok	95	(4,1%*)	129	[111, 175]
Nejasan u Visok	375	(18,6%*)	104	[76, 118]
Nizak u Nejasan	1144	(49,0%*)	56	[51, 63]
Prema nižem RoB	**214**	**(3,1%)**		
Visok u Nejasan	183	(7,1%*)	16	[15, 22]
Nejasan u Nizak	26	(1,3%*)	170	[104, 289]
Visok u Nizak	5	(0,2%*)	169‡	[60, 278]
Bez promjene	**5090**	**(73,6%)**		
Visok	2386	(92,7%*)	71	[67, 73]
Nejasan	1610	(80,1%*)	39	[33, 43]
Nizak	1094	(46,9%*)	129	[129, 135]

**postotak dodijeljenih procjena po podskupinama; †duljina komentara; ‡normalna raspodjela, aritmetička sredina*
Kratice: CI = interval pouzdanosti; RoB=rizik od pristranosti

Autori CSR, u RoB domeni o zasljepljivanju ispitanika i osoblja, najčešće uz ime domene naznače ishode zbog kojih je domena procijenjena; međutim u 22,1% analiziranih domena autori CSR nisu naveli koja je domena ishoda. U 73,1% analiziranih procjena, autori CSR precizirali su da se RoB domena odnosi na "sve ishode" (Tablica 10)

Tablica 10 – Raspodjela ishoda prema kategoriji i našoj ponovnoj procjeni

			Izračunata procjena rizika od pristranosti		
			Visok	**Nizak**	**Nejasan**
Kategorija ishoda	**N**	**%**	**N**	**N**	**N**
Svi ishodi	5054	73,1%	1998	809	2247
Nije pojašnjeno	1527	22,1%	644	222	661
Objektivni	96	1,4%	48	42	6
Subjektivni	78	1,1%	59	4	15
Komplikacije/nuspojave	53	0,8%	43	4	6
Ishodi osim mortaliteta	38	0,5%	24	14	0
Mortalitet	32	0,5%	5	27	0
Samo-procjena pacijenta	17	0,2%	15	1	1
Bihevioralni ishodi	13	0,2%	13	0	0
Procjena kliničara	10	0,1%	8	2	0
Ukupno	**6918**	**100,00%**	**2857**	**1125**	**2936**

2.2. Nesigurnost u popratnim komentarima

Najveća podskupina (Tablica 8) procjena koja nije bila u skladu s Cochraneovim priručnikom bila je ona u kojoj su autori CSR dodijelili nisku procjenu RoB, dok je naša procjena pokazala nejasan rizik od pristranosti (N = 1143, 62,6%). To je bilo zbog velike nesigurnosti (98,3% slučajeva) oko toga jesu li i ispitanici i osoblje bili zaslijepljeni. U 84,6% slučajeva ta nesigurnost dolazi iz dviju izjava: *double-blind* i placebo.

2.3. Utjecaj duljine popratnog komentara

Pokazali smo da LOC koji podupire procjenu utječe na dodijeljenu procjenu, postizanje zasljepljivanja i, također, na ponovnu procjenu (Tablica 11). Najkraći komentari s medijanom od 44 znaka pronađeni su u RoB tablicama s RCT-ima kojima je dodijeljen nejasan RoB za ovu domenu. Što je duži komentar (LOC veći od 90) - to je veća šansa da je bilo opisano uspješno zasljepljivanje, zajedno s niskom procjenom RoB za domenu o zasljepljivanju ispitanika i osoblja koju su dodijelili autori CSR ili je ponovna procjena u skladu s Cochraneovim priručnikom pokazala također niski RoB. Popratni komentari s LOC između 62 i 82 obično podržavaju procjenu visokog RoB ili navode neuspješno zasljepljivanje.

Tablica 11 – Utjecaj duljine popratnog komentara (LOC) na dodijeljenu procjenu, ponovnu procjenu i postizanje zasljepljivanja

		N (%)	**Medijan duljine komentara (znakovi)**	**95% CI**
RoB procjena iz Cochraneovih pregleda	Visok	2574 (37,2%)	66	[62, 69]
	Nejasan	2011 (29,0%)	44	[44, 48]
	Nizak	2333 (33,8%)	95	[90, 100]
RoB procjena iz ovog istraživanja	Visok	2856 (41,3%)	73	[72, 75]
	Nejasan	2937 (42,4%)	43	[40, 44]
	Nizak	1125 (16,3%)	129	[129, 136]
Zasljepljivanje posignuto	Visok	3118 (45,1%)	78	[75, 82]
	Nejasan	3008 (43,5%)	44	[42, 45]
	Nizak	792 (11,4%)	142	[133, 149]

Kratice: CI = interval pouzdanosti; RoB=rizik od pristranosti

2.4. Razdvajanje RoB domene na prema ishodima

Među svim analiziranim RCT-ima pronašli smo 237 RCT-a s više od jedne RoB procjene za domenu o zasljepljivanju ispitanika i osoblja, gdje je doneseno više procjena za različite ishode. U tih 237 RCT-a pronašli smo ukupno 609 procjena za različite ishode. Raspodjela procjena u ovoj podskupini značajno se razlikovala od glavnog uzorka (p <0,001) – vidi Tablica 12, s 61,7% (naspram 39,3%) procjena visokog RoB, 20,9% (naspram 15,8%) niskog i 17,4% (u odnosu na 44,8%) procjena nejasnog RoB u domeni o zasljepljivanju ispitanika i osoblja.

Tablica 12 – Razlike u raspodjeli procjena RoB prema različitim ishodima

	RoB procjena objavljena u Cochraneovim pregledima			**Nova RoB procjena izračunata prema Cochraneovom priručniku**		
Ishod	jedna	podijeljena	zajednička	jedna	podijeljena	zajednička
Procjena rizika	**N (%)**			**N (%)**		
visok	2253 (35,8%)	**319 (52,5%)**	2573 (37,2%)	2479 (39,3%)	**375 (61,7%)**	2856 (41,3%)
nizak	2147 (34,1%)	**185 (30,4%)**	2333 (33,7%)	998 (15,8%)	**127 (20,9%)**	1125 (16,3%)
nejasan	1902 (30,2%)	**104 (17,1%)**	2007 (29,%)	2825 (44,8%)	**106 (17,4%)**	2932 (42,4%)
Ukupno	6302 (100%)	608 (100%)	6913 (100%)	6302 (100%)	608 (100%)	6913 (100%)

U 139 RCT-a, sve procjene RoB s obzirom na različite ishode bile su jednake, dok smo u 98 RCT-a pronašli barem jednu različitu procjenu RoB između pod-domena. Naša ponovljena procjena, temeljena na Cochraneovom priručniku, rezultirala je sa 77 RCT-a (175 procjena) koji nisu imali jednoglasne procjene. Od 127 domena koje su procijenjene da imaju nizak RoB, kako od strane autora CSR, tako i od nas, u 69 (54,3%) je to bilo zbog popratnog komentara da je malo vjerojatno da izostanak zasljepljivanja ispitanika i osoblja može utjecati na ishod.

2.5. Razdvajanje domene prema zaslijepljenim subjektima – pod-domene za ispitanike i osoblje

U 605 analiziranih RCT-a, autori CSR nisu koristili domenu o zasljepljivanju ispitanika i osoblja, definiranu kao jedinstvenu, već su umjesto toga, RoB domenu podijelili na pod-domene na temelju dvije skupine pojedinaca: jedna pod-domena za zasljepljivanje ispitanika i/ili jedna pod-domena za zasljepljivanje osoblja. Međutim, nisu sva istraživanja koristila obje navedene pod-domene: 40 RCT-a imalo je samo pod-domenu za zasljepljivanje ispitanika, a

114 RCT-a imalo je samo pod-domenu za zasljepljivanje osoblja. U 451 RCT-u korištene su obje pod-domene sa ukupno 460 procjena RoB za zasljepljivanje ispitanika (veći je broj zbog 9 pokusa s daljnjom podjelom pod-domene na temelju ishoda) i 451 procjena za zasljepljivanje osoblja. Pronašli smo značajno niži udio nejasnih procjena i u pod-domeni za osoblje (22,7%, p <0,01) i u pod-domeni za ispitanike (18,4%, p <0,001), za razliku od 29,1% u glavnom uzorku (Tablica 13).

Tablica 13 – Razlike u raspodjeli procjena RoB između skupine kada je domena razdvojena prema subjektima i glavnog skupa podataka

Procjena Cochraneovih autora	**Nova procjena prema Priručniku**	**Ispitanici**	**Osoblje**	**Zajednički razdvojene domene**	**Glavni skup podataka**
		N (%)	**N (%)**	**N (%)**	**N (%)**
Visok rizik		**237 (47,4%)**	**319 (56,5%)**	**556 (52,2%)**	**2574 (37,2%)**
	Visok rizik	227 (95,8%)	307 (96,2%)	534 (96,0%)	2386 (92,7%)
	Nizak rizik	0 (0%)	0 (0%)	0 (0%)	5 (0,2%)
	Nejasan rizik	10 (4,2%)	12 (3,8%)	22 (4,0%)	183 (7,1%)
Nizak rizik		**171 (34,2%)**	**118 (20,9%)**	**289 (27,1%)**	**2333 (33,7%)**
	Visok rizik	4 (2,3%)	3 (2,5%)	7 (2,4%)	95 (4,1%)
	Nizak rizik	152 (88,9%)	105 (89,0%)	257 (88,9%)	1094 (46,9%)
	Nejasan rizik	15 (8,8%)	10 (8,5%)	25 (8,7%)	1144 (49,0%)
Nejasan rizik		**92 (18,4%)†**	**128 (22,7%)†**	**220 (20,7%)†**	**2011 (29,1%)†**
	Visok rizik	10 (10,9%)	6 (4,7%)	16 (7,3%)	375 (18,7%)
	Nizak rizik	6 (6,5%)	2 (1,6%)	8 (3,6%)	26 (1,3%)
	Nejasan rizik	76 (82,6%)	120 (93,8%)	196 (89,1%)	1610 (80,0%)
Podudarne procjene		**455 (91,0 %)**	**532 (94,2%)**	**987 (92,7%)**	**5090 (73,6%)**
Ukupno		**500 (100,0%)**	**565 (100,0%)**	**1065 (100,0%)**	**6918 (100,0%)**

** N/A = nije dostupno; † p < 0,05;*

Kada smo kombinirali procjene iz ove dvije pod-domene, postojalo je 92,7% podudaranja u kategorijama procjena (visokoj, niskoj, nejasnoj) koje su dodijelili autori CSR i onima koje smo ponovno procijenili (Tablica 13). To je bilo statistički značajno više podudarnih procjena u usporedbi s učestalošću podudarnih procjena za glavni skup podataka, gdje je podudaranje u procjenama 73,6%.

Nije bilo statistički značajne razlike između dodijeljenih i ponovno urađenih procjena u pod-domeni za zasljepljivanje osoblja (p = 0,2565, Tablica 23, dodatni materijali).

Kada je RoB procijenjen visokim, ukupno podudaranje u dvije skupine procjena bilo je 96,0% (Tablica 13); kada je procjena bila nizak RoB, podudarnost je bila 88,9%. Kada je RoB

procijenjen kao nejasan, slična je razina podudaranja pronađena samo u pod-domeni za zasljepljivanje osoblja, s 93,8% rezultata; u pod-domeni za zasljepljivanje ispitanika, podudaranje je bilo 82,6%. Ova je razlika bila značajna ($p < 0,05$), ali uzorak je bio vrlo malen (Tablica 23, dodatni materijali).

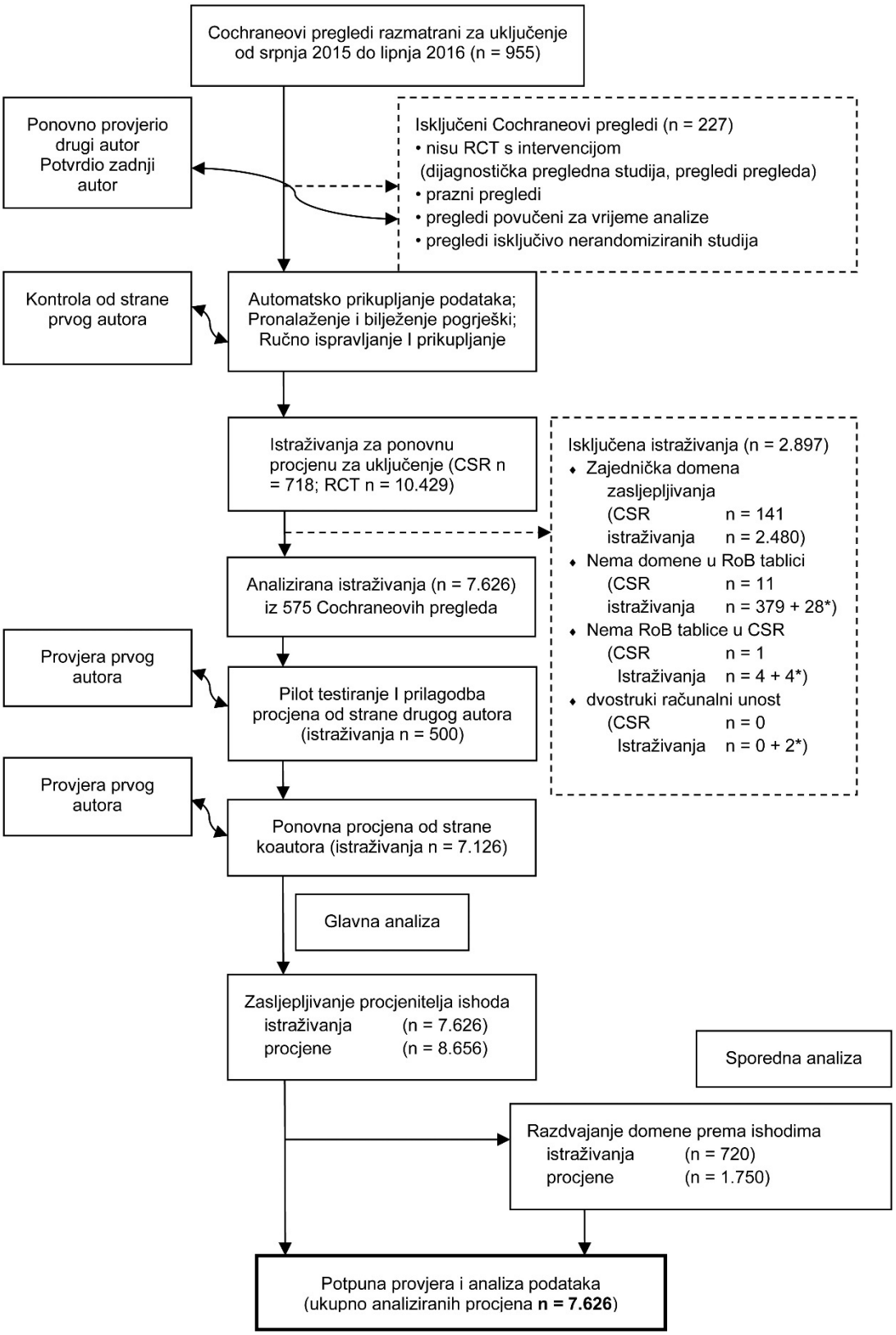
Pronašli smo ukupno 451 RCT istih CSR s procjenama RoB koje su uključivale obje pod-domene prema subjektima zasljepljivanja. U samo 82 RCT-a postojala je razlika u visine procjene RoB između pod-domena (ukupno 170 procjena RoB). Kvalitativno, pod-domena osoblja imala je tendenciju procjena višeg stupnja RoB i to s 28 visok RoB, 47 nejasan i 7 niskog RoB; dok je pod-domena o zasljepljivanju ispitanika pokazala tendenciju ukupno nešto niže procjene RoB: 12 visok, 17 nejasan i 59 nizak RoB.

3. Rezultati trećeg rada – prikladnost procjena Cochraneovih autora za domenu rizika pristranosti o zasljepljivanju osoba koje bilježe ishode

3.1. Uključeni CSR, RCT, procjene i ishodi

Od 955 obrađenih CSR, 227 nije ispunjavalo uvjete za uključenje. U preostalih 728 CSR bilo je 10523 RCT-ova. Iz analize je isključeno 2897 RCT-a: svih 2480 RCT iz 141 CSR u kojima su spojene domena o zasljepljivanju ispitanika i osoblja s domenom o zasljepljivanju osoba koje bilježe ishode; svih 379 RCT iz 11 CSR te dodatnih 28 RCT bez domene o zasljepljivanju osoba koje bilježe ishode navedene u RoB tablici; 1 sustavni pregled s 4 RCT-a bez RoB tablice i dodatna 4 RCT bez RoB tablice; 2 RCT bila su dvostruka unosa zbog računalne pogreške (Slika 5).

Konačno, u naše smo istraživanje uključili RoB tablice iz 575 CSR, koje su uključivale ukupno 7626 RCT (Slika 5). U tih 7626 RCT bilo je 8656 procjena za RoB o zasljepljivanju osoba koje bilježe ishode, jer je u nekim Cochraneovim pregledima ta domena bila razdvojena (imala je više procjena za različite vrste ishoda). U 720 od 7626 (9,4%) RCT, RoB domena o zasljepljivanju osoba koje bilježe ishode razdvojena je na više pod-domena na temelju različitih ishoda (u rasponu od 2 do 8). U tih 720 RCT-a bilo je 1750 procjena za određene ishode.



Slika 5 – Dijagram tijeka uključivanja i isključivanja istraživanja

U cijelom uzorku od 8656 procjena, za većinu (7110/8656; 82%) autori CSR naveli su u RoB tablici da se procjena odnosi ili na sve ishode (6072/8656; 70%) ili nisu naveli na koje se ishode odnosila domena (1038/8656; 12%). Za preostalih 18% procjena autori CSR naveli su na koje se ishode odnosi domena (ili pod-domena) - Tablica 14.

Tablica 14 – Raspodjela procjena RoB od strane autora CSR prema kategorijama ishoda i našoj ponovljenoj procjeni procjeni

Kategorija ishoda Procjena Cochraneovih autora	**Nova procjena u ovom istraživanju**							
	Visok rizik		**Nizak rizik**		**Nejasan rizik**		**Ukupno**	
Svi ishodi	**1116**	**(18%)**	**1255**	**(21%)**	**3701**	**(61%)**	**6072**	**(70%)**
Visok rizik	908	(81%)	5	(0%)	180	(5%)	1093	(18%)
Nizak rizik	41	(4%)	1208	(96%)	977	(26%)	2226	(37%)
Nejasan rizik	167	(15%)	42	(3%)	2544	(69%)	2753	(45%)
Procjena kliničara	**144**	**(22%)**	**185**	**(28%)**	**340**	**(51%)**	**669**	**(8%)**
Visok rizik	130	(90%)	1	(1%)	2	(1%)	133	(20%)
Nizak rizik	1	(1%)	182	(98%)	85	(25%)	268	(40%)
Nejasan rizik	13	(9%)	2	(1%)	253	(74%)	268	(40%)
Nije precizirano	**239**	**(23%)**	**247**	**(24%)**	**552**	**(53%)**	**1038**	**(12%)**
Visok rizik	221	(92%)	2	(1%)	35	(6%)	258	(25%)
Nizak rizik	2	(1%)	240	(97%)	172	(31%)	414	(40%)
Nejasan rizik	16	(7%)	5	(2%)	345	(63%)	366	(35%)
Objektivno mjereni	**0**	**(0%)**	**408**	**(100%)**	**0**	**(0%)**	**408**	**(5%)**
Visok rizik	0	/	9	(2%)	0	/	9	(2%)
Nizak rizik	0	/	292	(72%)	0	/	292	(72%)
Nejasan rizik	0	/	107	(26%)	0	/	107	(26%)
Samo-procjena pacijenata	**154**	**(72%)**	**55**	**(26%)**	**5**	**(2%)**	**214**	**(2%)**
Visok rizik	131	(85%)	5	(9%)	0	(0%)	136	(64%)
Nizak rizik	1	(1%)	50	(91%)	0	(0%)	51	(24%)
Nejasan rizik	22	(14%)	0	(0%)	5	(100%)	27	(13%)
Subjektivni	**59**	**(23%)**	**116**	**(45%)**	**80**	**(31%)**	**255**	**(3%)**
Visok rizik	50	(85%)	0	(0%)	0	(0%)	50	(20%)
Nizak rizik	1	(2%)	115	(99%)	7	(9%)	123	(48%)
Nejasan rizik	8	(14%)	1	(1%)	73	(91%)	82	(32%)
Ukupno	**1712**	**(20%)**	**2266**	**(26%)**	**4678**	**(54%)**	**8656**	**(100%)**

Raspodjela šest kategorija ishoda u cijelom uzorku (8656 procjena) i pod-uzorku s razdvojenom domenom (N = 720 RCT, N = 1750 procjena) bila je značajno različita ($p < 0,05$, Tablica 23, odjeljak dodatni materijali). U ovom je pod-uzorku utvrđeno samo 4% procjena za „sve ishode“, a 14% nije bilo precizirano (Tablica 15). U 231 RCT s razdvojenom domenom (koja je obuhvatila 540 procjena) procjene RoB bile su identične u svim njihovim podijeljenim

ishodima (primjerice, u jednom RCT sve procjene RoB-a bile su iste razine: sve visoke, sve niske ili sve nejasne).

Tablica 15 – Raspodjela i podudarnost procjena RoB prema ishodima

	Procjena za domene koje su podijeljene po ishodima			
	Ukupno	**Različito**	**Sve isto**	**Ukupan uzorak**
Vrsta ishoda	**N (%) - Adeq**	**N (%) - Adeq**	**N (%) - Adeq**	**N (%) - Adeq**
Svi	77 (4%) - 68%	48 (4%) - 52%	29 (5%) - 93%	6072 (70%) - 77%
Procjena kliničara	657 (38%) - 84%	427 (35%) - 82%	230 (43%) - 89%	669 (8%) - 84%
Nije precizirano	247 (14%) - 79%	166 (14%) - 71%	81 (15%) - 95%	1038 (12%) - 78%
Objektivni	407 (23%) - 71%	317 (26%) - 64%	90 (17%) - 97%	408 (5%) - 72%
Samo-procjena pacijenata	173 (10%) - 84%	117 (10%) - 83%	56 (10%) - 88%	214 (2%) - 87%
Subjektivni	189 (11%) - 92%	135 (11%) - 89%	54 (10%) - 98%	255 (3%) - 93%
Ukupno	**1750 – 81%**	**1210 – 75%**	**540 – 92%**	**8656 – 78%**

Adeq: Podudarnost

3.2. Podudarnost procjena RoB autora CSR u domeni o zasljepljivanju osoba koje bilježe ishode

U glavnoj analizi, među 8656 procjena RoB u domeni od zasljepljivanju osoba koje bilježe ishode, bilo je 1679 (19%) koje su autori CSR označili visokim rizikom, 3374 (39%) niskim rizikom i 3603 (42%) nejasnim rizikom od pristranosti (Tablica 16).

Tablica 16 – Podudarnost procjena RoB autora CSR s našom ponovnom procjenom za domenu o zasljepljivanju osoba koje bilježe ishode

Procjena Cochraneovih autora	**Nova procjena u ovom istraživanju**			**Ukupno N (%)**	**Neprikladne procjene N (%)**
	Visok rizik	**Nizak rizik**	**Nejasan rizik**		
Visok rizik	1440	**22**	**217**	**1679 (19,4%)**	**239 (14,2%)**
Nizak rizik	46	**2087**	**1241**	**3374 (39,0%)**	**1287 (38,1%)**
Nejasan rizik	226	**157**	**3220**	**3603 (41,6%)**	**383 (10,6%)**
Ukupno	**1712 (19,8%)**	**2266 (26,2%)**	**4678 (54,0%)**	**8656 (100,0%)**	**1909 (22,1%)**

Od 8656 procjena RoB, 6747 (78%) su autori CSR adekvatno označili. Naša ocjena primjerenosti tih procjena pokazala je da je najveća pojavnost nepodudarnih procjena utvrđena za RCT-ove označene niskim RoB za zasljepljivanje osoba koje bilježe ishode (1287 od 3374; 38%), a slijede ona označena visokim rizikom (239 od 1679; 14%) i potom ona s nejasnim rizikom (383 od 3603; 11%) (Tablica 16).

Podudarnost procjena za „sve ishode“ i procjena kod kojih nisu precizirani ishodi bila je 77%, odnosno 78% (Tablica 15 - desni stupac). Među preostalih 18% procjena, opažena je veća podudarnost među ishodima koje procjenjuje kliničar (84%), procjena / samo-procjena pacijenta (87%) i subjektivni ishodi (93%) nasuprot 78% u cijelom uzorku (svi $p < 0,05$, Tablica 24, odjeljak dodatni materijali). Ove tri skupine ishoda smatraju se subjektivnim i primijetili smo značajno veću podudarnost procjena u zajedničkoj skupini ishoda u usporedbi s cijelim uzorkom (86,9% naspram 77,9%, $p < 0,05$, Tablica 24, odjeljak dodatni materijali), kao i u usporedbi ove zajedničke skupine subjektivnih ishoda s podskupinom s podijeljenom domenom (85,7% vs 80,6%, $p < 0,05$). Konačno, procjene o objektivnim ishodima imale su nižu podudarnost (72%) u usporedbi s cijelim uzorkom (78%) ($p < 0,05$, Tablica 24, odjeljak dodatni materijali).

Pod-uzorak RCT-ova s razdvojenom domenom gdje su sve procjene RoB bile iste razine (Tablica 15) pokazao je veću podudarnost procjena (92% naspram 78% u cijelom uzorku $p < 0,05$). S druge strane, procjene u pod-uzorku RCT-a kad je domena razdvojena prema različitim ishodima i s različitim procjenama RoB, bili su jednako podudarni kao i u cijelom uzorku (76% naspram 78%, $p = 0,053$).

Duljina pratećeg komentara (LOC) koji podupire procjenu korelira s razinom procjene RoB od strane autora CSR, uspješnog zasljepljivanja i razinom naše ponovne procjene RoB u ovoj domeni. Najkraći komentari s medijanom od 45 znakova pronađeni su u RoB tablicama istraživanja s dodijeljenim nejasnim RoB (Tablica 17). Što je komentar duži (LOC preko 91), to je veća šansa da se radi o opisu uspješnog zasljepljivanja, zajedno s procjenom niskog RoB koju su dodijelili autori CSR ili smo mi ponovno procijenili prema Cochrane priručniku.

Tablica 17 – Utjecaj duljine popratnog komentara (LOC) na dodijeljenu procjenu, ponovnu procjenu i postizanje zasljepljivanja

		N (%)	Medijan duljine komentara (znakovi)	95% CI
Procjena RoB Cochraneovih autora	Visok	1712 (19,8%)	78	[75, 85]
	Nejasan	4678 (54,0%)	45	[44, 45]
	Nizak	2266 (26,2%)	102	[97, 108]
Procjena RoB izračunata u ovom istraživanju	Visok	1679 (19,4%)	70	[64, 74]
	Nejasan	3603 (39,0%)	43	[40, 44]
	Nizak	3374 (39,0%)	94	[91, 97]
Zasljepljivanje postignuto	Ne	1710 (19,8%)	79	[76, 85]
	Nejasno	5436 (62,8%)	48	[46, 49]
	Da	1510 (17,4%)	106	[101, 113]

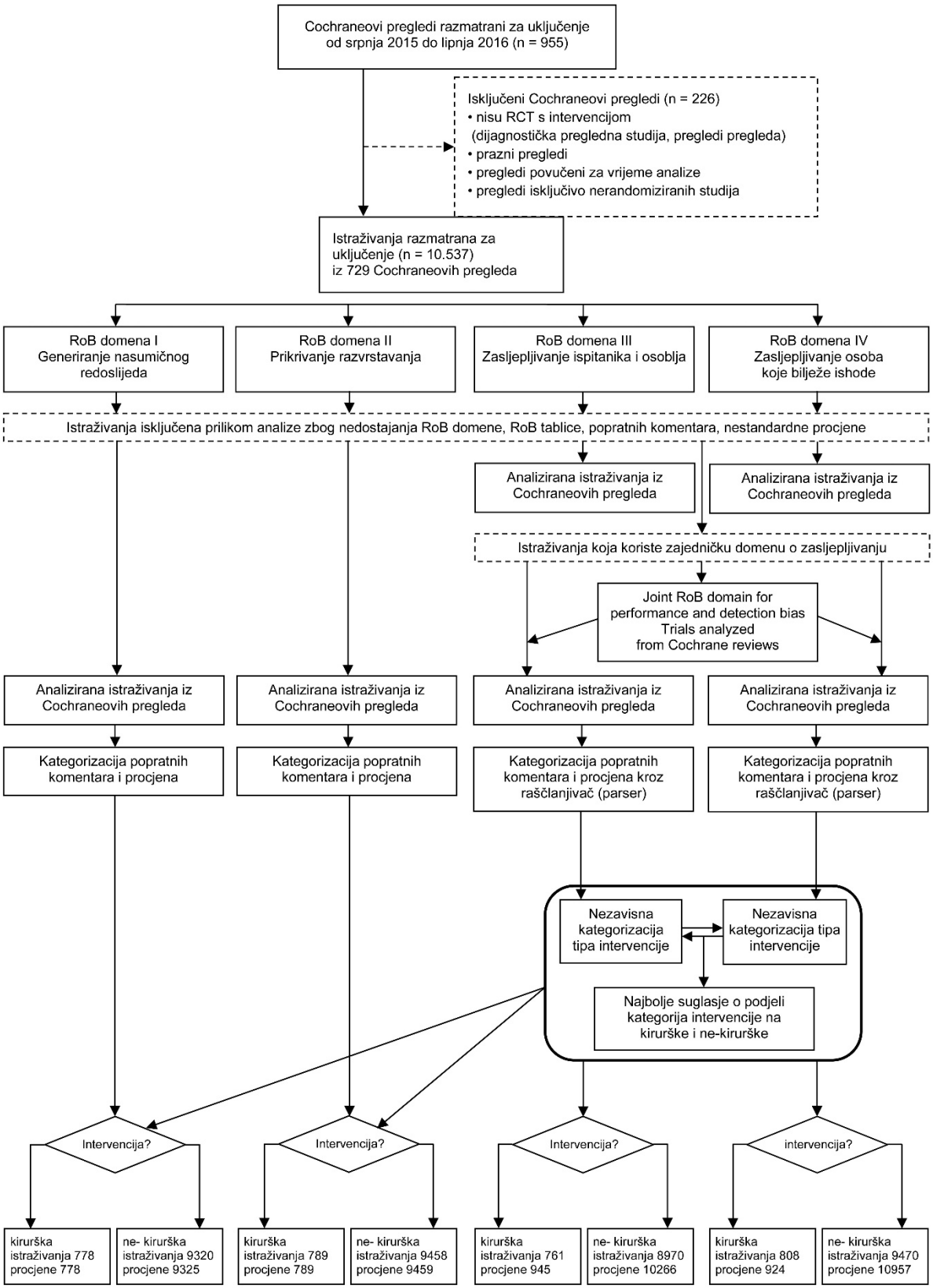
Kratice: CI = interval pouzdanosti; RoB=rizik od pristranosti

4. Rezultati četvrtog rada - istraživanje o kirurškim intervencijama

Analizirali smo 729 CSR, s 10537 uključenih istraživanja. Dijagram tijeka prikazan je na Slika 6. Nisu svi pregledi analizirali svih sedam standardnih RoB domena; domena generiranja nasumičnog redoslijeda analizirana je u pokusima iz 709 CSR, domena prikrivanja razvrstavanja iz 717 CSR, domena o zasljepljivanju ispitanika i osoblja iz 685 CSR te domena o zasljepljivanju osoba koje bilježe ishode iz 721 CSR (*Tablica 18*, Slika 6). U 171 analiziranom CSR, autori CSR koristili su zajedničku (jedinstvenu) domenu za procjenu zaslijepljenosti ispitanika, osoblja i osoba koje bilježe ishode.

Tablica 18 – Broj i udio Cochraneovih sustavnih pregleda, pokusi kojima nedostaju podatci za specifične domene, analizirani pokusi i procjene – ukupno i prema vrstama intervencija

		Domena I (generiranje nasumičnog redoslijeda)	**Domena II** (prikrivanje razvrstavanja)	**Domena III+*** (zasljepljivanje ispitanika i osoblja)	**Domena IV+*** (zasljepljivanje osoba koje bilježe ishode)
Uključeni pregledi (od izvornih 729 pregleda)	**Kirurški pokusi N (% od uključenih)**	77 (10,9%)	79 (10,7%)	72 (10,5%)	79 (11,0%)
	Ne-kirurški pokusi N (% od uključenih)	652 (92,0%)	659 (89,3%)	613 (89,5%)	642 (89,0%)
	Ukupno N (%)	709 (97,3%)	717 (98,4%)	685 (94,0%)	721 (98,9%)
Pokusi za koje nedostaju domene (od izvornih 10537 pokusa)	**Kirurški N (%)**	37 (0,4%)	28 (0,3%)	54 (0,5%)	16 (0,2%)
	Ne-kirurški N (%)	402 (3,8%)	329 (3,1%)	752 (7,1%)	417 (4,0%)
	Ukupno N (%)	439 (4,2%)	357 (3,4%)	806 (7,6%)	433 (4,1%)
Analizirani pokusi (od izvornih 10537 pokusa)	**Kirurški N (%)**	778 (7,4%)	789 (7,5%)	761 (7,2%)	808 (7,7%)
	Ne-kirurški N (%)	9320 (88,5%)	9391 (89,1%)	8970 (85,1%)	9296 (88,2%)
	Ukupno N (%)	10098 (95,8%)	10180 (96,6%)	9731 (92,4%)	10104 (95,9%)
Analizirane procjene (od broja analiziranih pokusa u domeni)	**Kirurški N (%)**	778 (7,7%)	789 (7,7%)	945 (8,4%)	924 (7,8%)
	Ne-kirurški N (%)	9325 (92,3%)	9459 (92,3%)	10266 (91,6%)	10957 (92,2%)
	Kirurški +N (+%)†	0 (+0,0%)	0 (+0,0%)	184 (+19,5%)	116 (+12,6%)
	Ne-kirurški +N (+%)†	5 (+0,1%)	68 (+0,7%)	1296 (+12,6%)	1661 (+15,2%)
	Ukupno N (+%)†	10103 (+0,0%)	10248 (+0,7%)	11211 (+13,2%)	11881 (+15,0%)



Slika 6 – Dijagram tijeka istraživanja

Domena za zasljepljivanje ispitanika i osoblja bila je prisutna u znatno manje CSR (n = 685) u usporedbi s ostale tri analizirane domene (P <0,001, Tablica 18). Nije bilo razlike u učestalosti korištenja četiri analizirane RoB domene između kirurških i ne-kirurških pregleda (u rasponu od 10,5% do 11,0% i 89,0% do 92,0%).

Najveća varijabilnost otkrivena je u udjelu odsutnih analiziranih domena u istraživanjima ne-kirurških intervencija (P <0,001, Tablica 18). To se najbolje moglo primijetiti u domeni o zasljepljivanju ispitanika i osoblja, jer više od 7% slučajeva nedostaje toj domeni.

Broj procjena bio je veći u obje domene o zasljepljivanju u odnosu na domene u vezi s randomizacijom i prikrivanjem razvrstavanja zbog toga što su autori CSR davali više procjena RoB (za različite ishode). Dakle, ukupni broj analiziranih procjena premašio je ukupan broj promatranih istraživanja, ali to je bilo proporcionalno i u kirurškim i u ne-kirurškim skupinama (P = 0,129, Tablica 18).

4.1. Kategorizacija intervencija

U konačnoj kategorizaciji intervencija od strane dva neovisna procjenitelja, konsenzus za kategorizaciju pokusa na kirurški i ne-kirurški bio je gotovo savršen (Cohenov Kappa 0,83, 95% CI [0,81 do 0,85]); detalji u Tablica 19.

Tablica 19 – Konsenzus (suglasje) dvaju procjenitelja i varijabilnost različitih kategorizacija intervencija

Domena Kategorije	**Suglasje**	**Cohenov Kappa**	**95%CI**	**Cohenov Kappa: interpretacija**
Domena III+* (zasljepljivanje ispitanika i osoblja)				
SURG vs MIXED vs CONS	83,0%	0,5744	0,5570-0,5918	umjereno suglasje
SURG vs NON-SURG	96,8%	0,7691	0,7457-0,7925	značajno suglasje
CONS vs NON-CONS	85,5%	0,6117	0,5942-0,6292	značajno suglasje
MIXED vs NON-MIXED	83,7%	0,4305	0,4066-0,4544	umjereno suglasje
Domena IV+* (zasljepljivanje osoba koje bilježe ishode)				
SURG vs MIXED vs CONS	87,1%	0,6645	0,6488-0,6801	značajno suglasje
SURG vs NON-SURG	97,8%	0,8329	0,8128-0,8529	gotovo savršeno suglasje
CONS vs NON-CONS	88,9%	0,6930	0,6774-0,7086	značajno suglasje
MIXED vs NON-MIXED	87,5%	0,5475	0,5261-0,5690	umjereno suglasje

** standardna domena spojena sa zajedničkom domenom za zasljepljivanje sve tri skupine osoba; SURG = kirurška intervencija; MIXED = miješana ili nejasna intervencija; CONS = konzervativna ili ne-invazivna intervencija; NON-SURG = sve osim kirurških; NON-CONS = sve osim konzervativnih; NON-MIXED = sve osim miješanih*

4.2. Raspodjela i podudaranje procjena

Raspodjela razina RoB (visokog / niskog / nejasnog) koje su dodijelili autori CSR (Tablica 20) za kirurška i ne-kirurška istraživanja nije se razlikovala za domene o randomizaciji i prikrivanja razvrstavanja (P = 0,409, P = 0,964, respektivno), ali su se značajno razlikovale u dvije domene oko zasljepljivanja (P <0,001, Tablica 25, odjeljak dodatni materijali).

Raspodjela RoB procjena koje smo donijeli *de novo*, na temelju objašnjenja iz tablica RoB, značajno se razlikovala između kirurških i ne-kirurških istraživanja u domeni za randomizaciju i domeni za zasljepljivanje osoba koje bilježe ishode (P = 0,022, P < 0,001, Tablica 25, odjeljak dodatni materijali). Gotovo je dosegla razinu statističke razlike za domenu u pogledu prikrivanja razvrstavanja (P = 0,0710).

Učestalost podudarnih procjena značajno je varirala između četiri RoB domene (Kruskal-Wallisov test, P <0,001, Tablica 21). U cijelom uzorku analiziranih CSR, najveća učestalost podudarnih procjena Cochraneova autora utvrđena je u RoB domeni za randomizaciju (87,9%, 95% CI [87,3% do 88,6%]), nakon čega slijedi domena za zasljepljivanje osoba koje bilježe ishode (72,9%, 95% CI [72,0% do 73,7%]), prikrivanje razvrstavanja (71,9%, 95% CI [71,0% do 72,8%]) i zasljepljivanje ispitanika i osoblja (70,7%, 95% CI [69,8% do 71,5%]).

Tablica 20 – Raspodjela procjena autora CSR i naših ponovljenih procjena prema tipu intervencije (kirurška / ne-kirurška)

RoB domena	Naša ponovljena procjena RoB u istraživanjima s kirurškim intervencijama					Naša ponovljena procjena RoB u istraživanjima s ne-kirurškim intervencijama				
I – generiranje nasumičnog redoslijeda Procjene autora CSR	**Visoki rizik**	**Nizak rizik**	**Nejasan rizik**	**Ukupno**	**Podu-darnost**	**Visoki rizik**	**Nizak rizik**	**Nejasan rizik**	**Ukupno**	**Podu-darnost**
Visoki rizik	31		1	32 (4.1%)	96.9%	233	20	80	333 (3.6%)	70.0%
Nizak rizik		337	58	395 (50.8%)	85.3%	9	3631	916	4556 (48.9%)	79.7%
Nejasan rizik	4	17	330	351 (45.1%)	94.0%	28	87	4321	4436 (47.6%)	97.4%
Ukupno N (%)	**35 (4.5%)**	**354 (45.5%)**	**389 (50.0%)**	**778 (100.0%)**	**89.7%**	**270 (2.9%)**	**3738 (40.1%)**	**5317 (57.0%)**	**9325 (100.0%)**	**87.8%**
II – prikrivanje razvrstavanja Procjene autora CSR	**Visoki rizik**	**Nizak rizik**	**Nejasan rizik**	**Ukupno**	**Podu-darnost**	**Visoki rizik**	**Nizak rizik**	**Nejasan rizik**	**Ukupno**	**Podu-darnost**
Visoki rizik	34		9	43 (5.4%)	79.1%	353		159	512 (5.4%)	68.9%
Nizak rizik		83	187	270 (34.2%)	30.7%	17	746	2479	3242 (34.3%)	23.0%
Nejasan rizik			476	476 (60.3%)	100.0%	16	10	5679	5705 (60.3%)	99.5%
Ukupno N (%)	**34 (4.3%)**	**83 (10.5%)**	**672 (85.2%)**	**789 (100.0%)**	**75.2%**	**386 (4.1%)**	**756 (8.0%)**	**8317 (87.9%)**	**9459 (100.0%)**	**71.7%**
III+* - zasljepljivanje ispitanika i osoblja Procjene autora CSR	**Visoki rizik**	**Nizak rizik**	**Nejasan rizik**	**Ukupno**	**Podu-darnost**	**Visoki rizik**	**Nizak rizik**	**Nejasan rizik**	**Ukupno**	**Podu-darnost**
Visoki rizik	266	1	40	307 (32.5%)	86.6%	3392	17	350	3759 (36.6%)	90.2%
Nizak rizik	25	108	59	192 (20.3%)	56.3%	130	1578	2006	3714 (36.2%)	42.5%
Nejasan rizik	49	20	377	446 (47.2%)	84.5%	505	87	2201	2793 (27.2%)	78.8%
Ukupno N (%)	**340 (36.0%)**	**129 (13.7%)**	**476 (50.4%)**	**945 (100.0%)**	**79.5%**	**4027 (39.2%)**	**1682 (16.4%)**	**4557 (44.4%)**	**10266 (100.0%)**	**69.9%**
IV+* - zasljepljivanje osoba koje bilježe ishode Procjene autora CSR	**Visoki rizik**	**Nizak rizik**	**Nejasan rizik**	**Ukupno**	**Podu-darnost**	**Visoki rizik**	**Nizak rizik**	**Nejasan rizik**	**Ukupno**	**Podu-darnost**
Visoki rizik	174	1	44	219 (23.7%)	79.5%	2003	34	359	2396 (21.9%)	83.6%
Nizak rizik	7	115	70	192 (20.8%)	59.9%	88	2374	2002	4464 (40.7%)	53.2%
Nejasan rizik	33	20	460	513 (55.5%)	89.7%	356	212	3529	4097 (37.4%)	86.1%
Ukupno N (%)	**214 (23.2%)**	**136 (14.7%)**	**574 (62.1%)**	**924 (100.0%)**	**81.1%**	**2447 (22.3%)**	**2620 (23.9%)**	**5890 (53.8%)**	**10957 (100.0%)**	**72.2%**

**uključuje podatke iz zajedničke domene o zasljepljivanju ispitanika, osoblja i osoba koje bilježe ishode*

Tablica 21 – Pregled varijabilnosti učestalosti prikladnih RoB procjena u analiziranim RoB domenama, prema vrsti intervencija u analiziranim pokusima, sa statističkim testovima i uparenim usporedbama

Domena – vrsta intervencija u pokusu		**Prikladnost**	**Test**	**Post hoc upareni test usporedbe**
Broj faktora	**Domena – kirurška intervencija**		**Kruskal-Wallis, P < 0,001**	**Conover, *P* < 0,05 Različit od faktora**
#1	Randomizacija	698/778 (89,7%)	značajna razlika u udjelu prikladnih procjena među domenama	#2 #3 #4
#2	Prikrivanje razvrstavanja	593/789 (75,2%)		#1 #4
#3	Zasljepljivanje ispitanika i osoblja*	751/945 (79,5%)		#1
#4	Zasljepljivanje osoba koje bilježe ishode *	749/924 (81,1%)		#1 #2
Broj faktora	**Domena - ne-kirurška intervencija**		**Kruskal-Wallis, P < 0,001**	**Conover, *P* < 0,05 Različit od faktora**
#1	Randomizacija	8185/9325 (87,8%)	značajna razlika u udjelu prikladnih procjena među domenama	#2 #3 #4
#2	Prikrivanje razvrstavanja	6778/9459 (71,7%)		#1 #3
#3	Zasljepljivanje ispitanika i osoblja*	7171/10266 (69,9%)		#1 #2 #4
#4	Zasljepljivanje osoba koje bilježe ishode *	7906/10957 (72,2%)		#1 #3
Broj faktora	**Domena - ukupno**		**Kruskal-Wallis, P < 0,001**	**Conover, *P* < 0,05 Različit od faktora**
#1	Randomizacija	8883/10103 (87,9%)	značajna razlika u udjelu prikladnih procjena među domenama	#2 #3 #4
#2	Prikrivanje razvrstavanja	7371/10248 (71,9%)		#1
#3	Zasljepljivanje ispitanika i osoblja*	7922/11211 (70,7%)		#1 #4
#4	Zasljepljivanje osoba koje bilježe ishode *	8655/11881 (72,8%)		#1 #3

**uključuje podatke iz zajedničke domene o zasljepljivanju ispitanika, osoblja i osoba koje bilježe ishode*

Učestalost podudarnih procjena RoB-a za sve analizirane RoB domene općenito je bila veća u kirurškim pokusima nego u ne-kirurškim pokusima. Za dvije RoB domene koje procjenjuju zasljepljivanje, ova razlika između kirurških i ne-kirurških istraživanja bila je statistički značajna (P <0,001), za istraživanje u domeni prikrivanja razvrstavanja snaga testa je bila preniska (P = 0,04039, beta <0,8), dok između istraživanja za RoB domenu generiranja nasumičnog redoslijeda nije bilo značajne razlike s obzirom na vrstu intervencije (P = 0,1312) (Tablica 25, odjeljak dodatni materijali).¸

4.3. Osnova za obrazloženja procjena RoB

Razni komentari korišteni su za potporu procjenama RoB-a u CSR. U RoB domeni za randomizaciju pokazali smo značajno različitu raspodjelu kategorija različitih pratećih komentara u kirurškoj i ne-kirurškoj skupini (P <0,001, tablica 3, domena I). Za kirurška istraživanja češće je spominjano kompjutorizirano i neodgovarajuće randomiziranje, a neuspjeh opisivanja metode randomizacije bio je rjeđi u usporedbi s ne-kirurškim istraživanjima (Tablica 22, Tablica 26, odjeljak dodatni materijali).

U RoB domeni za prikrivanje razvrstavanja pronašli smo sličnu raspodjelu vrsta prikrivanja razvrstavanja između kirurških i ne-kirurških istraživanja (Tablica 22, Domena II). Međutim, u kirurškim pokusima (u odnosu na ne-kirurške) otkrili smo da je veći udio komentara koji navode da je prikrivanje razvrstavanja ispravno postignuto upotrebom "uzastopno numeriranih neprozirnih zatvorenih omotnica" (SNOSE) i nižim udjelom nejasno opisanih metoda prikrivanja razvrstavanja.

Obje RoB domene o zasljepljivanju imale su značajno različitu raspodjelu komentara o tome je li postignuto zasljepljivanje ključnih osoba između kirurških i ne-kirurških istraživanja (P <0,001, Tablica 22, domena III i IV). Uspješno zasljepljivanje bilo je značajno rjeđe u kirurškim nego ne-kirurškim istraživanjima, za ispitanike i osoblje (4,0% naspram 12,1%, P <0,001), i za osobe koje bilježe ishode (9,1% naspram 15,7%, P <0,001).

Neki su ishodi osjetljiviji na pristranost zbog nedostatka zasljepljivanja u usporedbi s drugima; međutim, osjetljivost ishoda na koje može utjecati nedostatak zasljepljivanja ponekad je opisana u manje od 10% komentara za obje domene o zasljepljivanju (Tablica 22). Analizirali smo postoji li razlika u raspodjeli komentara u koje su autori Cochranea uključili informacije ako je zasljepljivanje ključnih pojedinaca utjecalo na ishod. Ta se raspodjela značajno razlikovala između kirurških i ne-kirurških istraživanja samo u RoB domeni koja procjenjuje zasljepljivanje ispitanika i osoblja (P <0,001, Tablica 22, Tablica 26, odjeljak dodatni materijali).

Tablica 22 – Diferencijacija obrazloženja (uzroka) za procjenu rizika prema testovima i interpretacijama

	Opravdanje za procjenu rizika	**Intervencija**	
Domena I (randomizacija)	**Kategorizacija popratnog komentara**	**Ne-kirurška**	**Kirurška**
	Randomizacija nije opisana	5317 (57,0%)	389 (50,0%)
	Tablica nasumičnih brojeva	815 (8,7%)	68 (8,7%)
	Računalna randomizacija	2597 (27,8%)	253 (32,5%)
	Mehanička randomizacija	326 (3,5%)	33 (4,2%)
	Neprikladna randomizacija	270 (2,9%)	35 (4,5%)
Domena II (prikrivnanje razvrstavanja)	**Vrsta prikrivanja razvrstavanja**	**Ne-kirurška**	**Kirurška**
	Središnje razvrstavanje	506 (5,3%)	32 (4,1%)
	Nepotpuni opis kuverti	1117 (11,8%)	147 (18,6%)
	Nije opisano/nejasno	7200 (76,1%)	525 (66,5%)
	Otvoreno/predvidljivo razvrstavanje	386 (4,1%)	34 (4,3%)
	Prikladan opis kuverti (SNOSE)	250 (2,6%)	51 (6,5%)
Domena III+* (zasljepljivanje ispitanika i osoblja)	**Zasljepljivanje postignuto**	**Ne-kirurška**	**Kirurška**
	Postignuto	1239 (12,1%)	38 (4,0%)
	Nije učinjeno ili nije moguće	4318 (42,1%)	414 (43,8%)
	Vjerojatno postignuto	4709 (45,9%)	493 (52,2%)
	Ishod na koji utječe izostanak zasljepljivanja	**Ne-kirurška**	**Kirurška**
	Utječe	472 (4,6%)	21 (2,2%)
	Nepoznato	9340 (91,0%)	832 (88,0%)
	Ne utječe	454 (4,4%)	92 (9,7%)
Domena IV+* (zasljepljivanje osoba koje bilježe ishode)	**Zasljepljivanje postignuto**	**Ne-kirurška**	**Kirurška**
	Postignuto	1722 (15,7%)	84 (9,1%)
	Vjerojatno postignuto	6695 (61,1%)	609 (65,9%)
	Nije učinjeno ili nije moguće	2540 (23,2%)	231 (25,0%)
	Ishod na koji utječe izostanak zasljepljivanja	**Ne-kirurška**	**Kirurška**
	Utječe	919 (8,4%)	31 (3,4%)
	Nepoznato	9045 (82,5%)	839 (90,8%)
	Ne utječe	993 (9,1%)	54 (5,8%)

***** *također uključuje podatke za zajedničke dvije domene za zasljepljivanje; SNOSE = uzastopno numerirane neprozirne zatvorene omotnice (engl. sequentially numbered sealed opaque envelopes)*

RASPRAVA

1. Prikladnost procjena Cochraneovih autora za domenu rizika pristranosti o generiranju nasumičnog redoslijeda

Analiza procjena i popratnih komentara o RoB povezanog s generiranjem nasumičnog redoslijeda u 10103 RCT-a uključenih u 704 CSR-a pokazala je da se autori CSR često ne pridržavaju preporuka u vezi s procjenom ove određene vrste RoB, jer je 12% procjena za pristranost bilo pogrešno (nepodudarno). To znači da je jedna od svakih osam procjena RoB u vezi s nasumičnim redoslijedom ispitanika u Cochraneovim pregledima ili pogrešna ili nije dovoljno razrađena. Najčešće pogreške uočene su u kategoriji istraživanja za koja se procijenilo da imaju visok RoB, gdje 28% procjena nije bilo potkrijepljeno objašnjenjima u popratnim komentarima. Prema našim najboljim saznanjima, ovo je prvo istraživanje koje je procijenilo opravdanost i točnost procjena rizika od pristranosti povezanih s generiranjem nasumičnog redoslijeda u CSR-ovima.

Najčešća pogreška, prema definiranim kategorijama, bila je procjena RoB-a za generiranjem nasumičnog redoslijeda kao niska kada metoda randomizacije nije bila dovoljno opisana. Najčešći takvi popratni komentari odnosili su se na spominjanje središnje randomizacije bez dodatnih detalja, zatim na pojam ravnoteže na početku istraživanja (engl. *baseline balance or imbalance*) između skupina u kojima autori CSR pretpostavljaju da je generiranje nasumičnog redoslijeda bilo adekvatno, komentare o randomizaciji u blokovima ili stratifikaciji te popratne komentare o korištenju omotnica bez daljnjih pojedinosti.

Cochraneov priručnik izričito upozorava: *"Ponekad autori istraživanja daju neke informacije, ali oni nepotpuno definiraju svoj pristup i ne navedu nasumičnu komponentnu u procesu. Na primjer, autori mogu navesti da je korištena randomizacija u blokovima, ali postupak odabira blokova, poput tablice nasumičnih brojeva ili računalnog generatora nasumičnih brojeva, nije naveden. Ispravnost generiranja redoslijeda tada treba klasificirati kao nejasnu "* (60).

Cochrane trenutno provodi implementaciju revidiranog alata za procjenu RoB u randomiziranim istraživanjima (RoB 2) (78). U novoj verziji alata osnovne značajke grupe ispitanika su na istaknutom mjestu. RoB 2 koristi osnovnu neravnotežu kako bi signalizirao probleme s postupkom randomizacije, a to je i jedno od signalnih pitanja u RoB domeni o procesu randomizacije (78).

Otkrili smo da je spomenuto 98 slučajeva generiranja nasumičnih brojeva bez primjedbi o korištenju računala, elektroničkih kalkulatora itd. Ova je kategorija ocijenjena kao 100% nizak rizik, a ostavili smo je u kategoriji računalne randomizacije iako nije udovoljavala strogim pravilima Cochraneovog priručnika. Inače bi se pogrešne razine povisile s 12% na 13% ukupno i s 20% na 22% za istraživanja za koja se smatra da imaju nizak rizik od pristranosti.

Naši rezultati imaju dalekosežan učinak na propitivanje pouzdanosti zaključaka u CSR i brojnim meta-epidemiološkim istraživanjima koja su se temeljile na procjenama RoB-a iz CSR, a u kojima autori nisu propitivali prikladnost procjena Cochraneovih autora. Procjena RoB-a redovito se spominje u zaključcima CSR-ova. Na primjer, naši podaci pokazuju da je svaka četvrta procjena, koja ukazuje na to da istraživanje uključeno u CSR ima visok RoB povezanog s generiranjem nasumičnog redoslijeda, pogrešna.

U CSR vrlo se često može pronaći da su obuhvaćena istraživanja bila visokog RoB, pa su stoga njihovi rezultati manje pouzdani. S druge strane, ako se potraga za visokokvalitetnim istraživanjima temelji na istraživanjima s niskim RoB-om, gdje je svaka peta procjena pogrešna, tada su zaključci CSR-ova ozbiljno ugroženi. Čitatelji takvih sustavnih pregleda dobivaju poruku da se određene preporuke oslanjaju na dokaze s niskim RoB, a time i na istinite i pouzdane dokaze.

Nedavno istraživanje opisalo je da su se loše izvješćivanje i neadekvatne metode s vremenom smanjivale, posebno za generiranje nasumičnog redoslijeda i njegovo prikrivanje, na temelju analize 20920 RCT-a uključenih u CSR-ove (99). Međutim, naše je istraživanje pokazalo da taj rezultat ne mora biti posljedica boljeg izvješćivanja i boljih metoda, već pogreški u procjenama autora CSR.

U našem istraživanju otkrili smo da je više od polovice istraživanja uključenih u naš uzorak CSR-a imalo nejasan rizik od pristranosti za generiranje nasumičnog redoslijeda, što je u skladu s izvješćem Kahana i sur., koji su izvijestili da je rizik od pristranosti za randomizaciju teško utvrditi u većini istraživanja upravo zbog lošeg izvješćivanja (100).

Isto tako, istraživanja koja su analizirala vezu između RoB-a i učinka intervencija mogla su doći do pogrešnih zaključaka jer su vjerovala prosudbi autora sustavnih pregleda. Savović i sur. izvijestili su da su procjene učinka intervencije bile pretjerane u prosjeku za 11% u kliničkim istraživanjima s neadekvatnim ili nejasnim generiranjem nasumičnog redoslijeda. Njihovi rezultati temeljili su se na analizi 1973 istraživanja obuhvaćenih u 234 meta-analize (101).

Prema Cochraneovom priručniku, autori Cochranea mogu iznositi pretpostavke, ali ih trebaju detaljno objasniti (60). U ovom su istraživanju pronađeni popratni komentari u RoB tablicama, gdje su autori Cochranea upravo napisali pretpostavke i procijenili da je istraživanje

imalo nizak rizik od pristranosti povezan s randomizacijom, ali bez objašnjenja zašto pretpostavljaju da je rizik nizak.

Ograničenje ovog istraživanja je vremensko razdoblje u kojem su objavljeni analizirani CSR. Međutim, analizirali smo velik broj CSR-ova, s velikim brojem uključenih istraživanja i to u vrijeme kada je upotreba RoB 2011 alata bila u potpunosti implementirana. Stoga vjerujemo da su oni reprezentativni za trenutno stanje izvješćivanja o analiziranoj domeni u CSR. Sljedeće ograničenje je moguća ljudska pogreška zbog interpretacije Cochraneova priručnika i ručne kategorizacije automatski izdvojenog teksta. Da bi se spriječila ova pogreška, svaka je kategorizacija bila ponovno provjerena. Također su svi podatci javno objavljeni na OSF-u, kako bi sve procjene bile transparentne.

Već je pokazano da procjene učinjene Cochraneovim RoB alatom imaju nisku podudarnost između pojedinih autora sustavnih pregleda, kao i među-konsenzusnih procjena parova autora (102). Da Costa i sur. tvrdili su da niska pouzdanost procjene RoB-a u sustavnim pregledima može imati štetne učinke na donošenje odluka i kvalitetu zdravstvene zaštite (103). Testirane su intervencije poput standardiziranog intenzivnog treninga o procjeni RoB-a, a rezultati ukazuju da takve intervencije mogu značajno poboljšati pouzdanost Cochraneovog RoB alata (104). Osim edukacije autora, druga rješenja za poboljšanje pouzdanosti procjena RoB-a bila bi stroža stručna recenzija i urednička analiza procjena i popratnih komentara u Cochraneovoj RoB tablici.

Uz to, u ovom radu analizirali smo samo CSR koje koriste Cochraneov RoB alat. Nedavna istraživanja pokazala su da RoB alat često ne prikaže na odgovarajući način rizik od pristranosti u RCT-ima (105, 106). Buduća istraživanja o alatu Cochrane RoB 2 opravdana su kako bi se vidjelo kako se novi alat uspoređuje sa sadašnjim po pitanju procjene rizika od pristranosti povezanog s generiranjem nasumičnog redoslijeda.

2. Prikladnost procjena Cochraneovih autora za domenu zasljepljivanja ispitanika i osoblja

Ovo je istraživanje otkrilo da u četvrtini analiziranih Cochraneovih pregleda procjena RoB zbog zasljepljivanja ispitanika i osoblja nije u skladu s preporukama iz Cochraneovog priručnika, ključnog izvora metodoloških smjernica za autore CSR.

Naši se rezultati mogu koristiti za poboljšanje budućih procjena RoB-a i za usmjeravanje daljnjih revizija RoB alata. Našli smo posebno velik broj odstupanja kada su Cochraneovi autori procijenili da je RoB nizak, ponajviše zato što su autori CSR prihvatili izjave kao što su „dvostruko slijepi“ (engl. *double-blind*) i „placebo“ kao dovoljan dokaz o uspjehu zasljepljivanja ispitanika i osoblja. Cochraneov priručnik posebno ukazuje na to da se pojam dvostruko slijepo koristi nedosljedno i onemogućava saznanje tko je bio zaslijepljen. Zabrinjavajuće je što, unatoč jasnom primjeru u uputama, mnogi autori CSR ove izraze povezuju s niskom procjenom RoB. Literatura sugerira sedam skupina ključnih pojedinaca koje mogu biti zaslijepljene tijekom istraživanja *(61, 72, 90, 91)* pa bi stoga autori sustavnih pregleda trebali jasno navesti jesu li u istraživanju jasno opisane informacije o određenim skupinama subjekata zasljepljivanja.

Nadalje, naši rezultati pokazuju da bi se mogle donijeti preciznije procjene ako je domena o zasljepljivanju ispitanika i osoblja podijeljena na dvije pod-domene, za ispitanike i za osoblje. U novijoj inačici Cochraneova alata za rizik od pristranosti za randomizirana istraživanja (RoB 2) koriste se signalna pitanja, a postoje zasebna signalna pitanja o zasljepljivanju ispitanika kao i osoblja (engl. *carers and people delivering the intervention*) (78). Rezultati našeg istraživanja podržavaju tu reviziju.

Otkrili smo da su autori CSR-ova prilagodili neke od analiziranih domena navodeći ishode na koje se domena odnosi. *Review Manager (RevMan)* (95), računalni program koji se koristi za izradu CSR, mogao bi se poboljšati kako bi naveo autore da obvezno označe je li se njihova procjena odnosi na sve ishode, subjektivne ili objektivne ishode, tako da u budućnosti neće biti slučajeva u kojima to nije određeno. Međutim, trebalo bi napustiti razdvajanje domene na temelju različitih ishoda.

Zbog naših rezultata koji pokazuju kako je kraća duljina pratećih komentara (LOC) povezana s većom učestalosti pogrešnih procjena, u *RevMan* bi se moglo uvesti minimalno ograničenje broja znakova, a autore upozoriti kad je komentar prekratak; granična vrijednost koju preporučujemo iznosi 44 znaka.

Osim intervencija na razini Cochrane RoB alata, mogu se uvesti i intervencije usmjerene na razinu urednika/recenzenata kako bi se pomoglo autorima CSR u pravilnoj procjeni RoB.

Ovo je istraživanje imalo nekoliko potencijalnih ograničenja. Iako smo se u ovom istraživanju koristili računalnim programima, ne možemo isključiti mogućnost ljudske pogreške u interpretaciji podataka. Nadalje, moguće je da su procjene RoB-a autora CSR, ustvari, ispravne, ali su se činile neadekvatnima zbog nedovoljnih detalja u popratnim komentarima; to bi se moglo provjeriti u uključenim RCT-ovima ili kontaktiranjem autora. U ovom istraživanju bili smo usmjereni samo na Cochraneove sustavne preglede jer je upotreba Cochraneovog RoB alata obavezna u Cochraneovim sustavnim pregledima. Ipak, naši rezultati mogu biti korisni i autorima ne-Cochraneovih sustavnih pregleda koji isto, u značajnoj mjeri, koriste Cochraneov RoB alat.

3. Prikladnost procjena Cochraneovih autora za domenu rizika pristranosti o zasljepljivanju osoba koje bilježe ishode

Glavni rezultat ovog istraživanja je da je 22% procjena RoB-a za domenu o zasljepljivanju osoba koje bilježe ishode u analiziranim CSR bilo neadekvatno, jer te procjene nisu potkrijepljene odgovarajućim popratnim komentarima. Pronašli smo podudarnije procjene u istraživanjima gdje su autori CSR razdvajali ovu RoB domenu prema različitim kategorijama ishoda, poput subjektivnih ili objektivnih ishoda.

Kada smo prethodno analizirali domenu za zasljepljivanje ispitanika i osoblja, zaključili smo da postoje četiri aspekta koja treba uzeti u obzir za donošenje procjene: tko je bio zaslijepljen; je li postignuto zasljepljivanje; kategorija ishoda; te može li na ishod utjecati izostanak zasljepljivanja (82).

Isti se principi mogu primijeniti za donošenje procjena RoB u domeni o zasljepljivanju osoba koje bilježe ishode. Prvo je pitanje tko su bile te osobe koje su bilježile ishode i jesu li jasno naznačene u izvještaju o istraživanju. Pokusi obično imaju više ishoda, pa stoga može postojati više osoba koje bilježe ishode. Drugo je pitanje jesu li osobe koje bilježe ishode bile zaslijepljene. Treće i četvrto pitanje su koje se vrste ishoda procjenjuju i može li se na rezultate tih ishoda utjecati nedostatkom zasljepljivanja osoba koje ih bilježe. Sve se ove informacije često ne prijavljuju u istraživanju ili se o njima slabo izvješćiva, kako su pokazali Kahan i suradnici (100).

Međutim, autori originalnih RCT nisu jedini s lošim opisima ishoda koji su ocjenjivani. Opisni pojmovi koje su autori CSR koristili za opisivanje ishoda koji su procijenjeni često nisu dovoljno specifični. Primjer je deskriptor „objektivno ishod“. U engleskom rječniku (107) pojam objektivan (engl. *objective*, kao pridjev) znači „nije pod utjecajem“, što sugerira strogi logički zaključak o niskom riziku od pristranosti. Ako se ishod pogrešno kategorizira kao objektivan, to automatski dovodi do pogrešne procjene. Jedan od takvih primjera je duljina boravka (108) - brojčana vrijednost koja se može smatrati objektivnom, ali zapravo jako ovisi o subjektivnim odlukama nadležnog liječnika. Moustgaard i sur. (94) posebno su raspravljali o ovom pitanju objektivnosti nasuprot kliničkoj važnosti. Isto je istraživanje navelo tri definicije pojma "subjektivno" i, prema njima, "objektivni" pojmovi definirani su kao suprotnosti.

Pokazali smo da autori CSR rijetko opisuju ishode kao subjektivne ili kao objektivne. Također nije dovoljno samo naznačiti da je ishod procijenio kliničar. Neki autori CSR opisuju ishod kao vezan za kliničara, ali s konkretnim primjerima takvog ishoda, koji pojašnjavaju prirodu ishoda i omogućuju lakšu prosudbu je li procjena ishoda sklona subjektivnosti. Izjava

da je ishod bio prijavljen od strane pacijenta ili samoprijavljen, implicira da je procjena bila subjektivna. Jasni opisi rezultata koji su procijenjeni mogu se ugraditi u novije verzije alata za procjenu RoB-a u obliku programiranih pravila.

Iako je središnji dio ovog istraživanja (i promatrane domene) bio naglasiti diferencijaciju ishoda kao specifičniji čimbenik, moramo spomenuti i kako duljina popratnog komentara (LOC) zamagljuje opravdanost procjene. Čak smo i u prethodnom istraživanju (82), o domeni o zasljepljivanju ispitanika i osoblja, pokazali da duljina komentara utječe na adekvatan (transparentan) opis uspješnog zasljepljivanja, a isto vrijedi i za domenu o zasljepljivanju osoba koje bilježe ishode.

3.1. Treba li RoB domenu o zasljepljivanju osoba koje bilježe ishode razdvajati prema ishodima, kojim ishodima i zašto?

Primijetili smo znatno veću podudarnost procjena RoB za ishode koji su opisani kao subjektivni, u usporedbi s drugim kategorijama ishoda, kako u RCT-ima s razdvojenom domenom, tako i u cijelom uzorku pokusa. Također smo pronašli inverziju u proporcijama podudarnih procjena, ovisno o tome je li domena o zasljepljivanju osoba koje bilježe ishode razdvojena na pod-domene ili ne. U pokusima u kojima domena nije bila razdvojena, udio podudarnih procjena za specificirane ishode u odnosu na one koji su ukazivali da su prosuđivali sve ishode ili nisu određivali vrstu ishoda bili su 18% u odnosu na 82%. Među istraživanjima koja su imala razdvojenu domenu za različite ishode, udio podudarnih procjena bio je obrnut - 82% kada su specificirane pod-kategorije ishoda u odnosu na 18% za sve ishode ili kada ishodi nisu navedeni. Nadalje, u podskupini istraživanja s razdvojenom domenom bila je veća pojavnost nepodudarnih procjena RoB za objektivne ishode nego za ostale skupine ishoda. Vjerujemo da je to zbog pogrešne kategorizacije ishoda poradi nedostatka specifičnih uputa u Cochraneovom priručniku o tome koje bi ishode trebalo smatrati objektivnima.

Čini se da razdvajanje procjene RoB u domeni o zasljepljivanju osoba koje bilježe ishode na više pod-domena ima neke specifičnosti. Prvo, ova je podskupina imala nešto veću podudarnost procjena u odnosu na cijeli uzorak, ali to je zbog većeg udjela ishoda opisanih kao subjektivnih. Subjektivni ishodi imaju veću podudarnost procjena, ali velik broj slučajeva opisanih kao objektivnih umanjuje učinak poboljšanja ukupne podudarnosti zbog redovito pogrešne procjene RoB za objektivne ishode. Drugo, autori često dodjeljuju istu razinu procjene RoB za sve vrste ishoda unutar jednog RCT, bez obzira na kategoriju ishoda. Treće, iako je to bilo rijetko, u 4% RCT s podijeljenim ishodima autori su također koristili pod-domenu "svi ishodi", zajedno s jednim ili više specifičnih ishoda. Ako su autori odlučili domenu podijeliti na više pod-domena

na temelju specificiranih ishoda, nejasno je zašto bi onda, dodatno, koristili pod-domene za 'sve ishode'.

Razdvajanje domene o zasljepljivanju osoba koje bilježe ishode može dovesti do krive procjene RoB, ali to je zbog pogreške u kategorizaciji (objektivnosti) ishoda. Međutim, ovakav prikaz je kritičan i dovodi na vidjelo pogrešnu kategorizaciju određenih ishoda, a čitateljima pruža više informacija nego zbirna procjena za sve ishode. Stoga možemo zaključiti da procjena RoB za sve ishode u ovoj domeni nije dobar izbor te da bi autori CSR barem trebali definirati kategoriju ishoda, pojasniti da svi ishodi pripadaju istoj kategoriji te da je to bio razlog zašto su svi zajedno procijenjeni.

Cochrane je nedavno objavio novu verziju svog RoB alata (78). U trenutku izrade ovog istraživanja (prosinac 2019.), i u trenutku predaje ove disertacije (siječanj 2020.), novi alat još nije bio u potpunosti implementiran u Cochraneovim sustavnim pregledima. Čak i autori Cochraneovih protokola koji još nisu objavljeni nisu bili obvezni koristiti RoB 2.

U RoB 2 promijenjen je sadržaj, struktura i vrsta primjenjivih procjena. RoB alat iz 2011 ima sedam domena, a svaka se domena može ocjenjivati s tri razine procjene – rizik se smatra visokim, niskim ili nejasnim. Alat RoB 2 ima pet domena; svaka domena ima od tri do sedam signalnih pitanja; na signalna pitanja odgovara pet potencijalnih odgovora (da, vjerojatno da, ne, vjerojatno ne ili nema podataka). Na temelju odgovora na signalna pitanja, svaka se domena ocjenjuje s RoB kao niskim, visokim ili da postoje neke nedoumice. Također je pružena cjelokupna ocjena RoB-a. Domena #4 je "Pristranost u mjerenju ishoda"; ova domena ima tri signalna pitanja koja se odnose na zasljepljivanje osoba koje bilježe ishode, postavljajući pitanje (i) jesu li osobe koje bilježe ishode bile zaslijepljene o tipu intervencije, (ii) može li na procjenu ishoda utjecati izostanak zasljepljivanja osoba koje biježe ishode i (iii) je li vjerojatno je da je na procjenu zapravo utjecao izostanak zasljepljivanja osoba koje biježe ishode. Ova tri signalna pitanja predstavljaju veliku promjenu u odnosu na RoB alat iz 2011. Vrlo slična pitanja, uz sličnu kategorizaciju mogućih odgovora, i s istim logičkim slijedom korištena su u dva naša istraživanja o zasljepljivanju (82, 87) tijekom ponovne procjene u korisničkom sučelju našeg parsera. Na temelju rezultata ovog istraživanja, očekujemo da bi ova tri signalna pitanja trebala olakšati procjene RoB za domenu o zasljepljivanju osoba koje bilježe ishode i trebala bi pomoći u donošenju transparentnih, adekvatno objašnjenih odluka.

Rezultati našeg istraživanja mogu pomoći u poboljšanju prikladnosti procjene RoB-a u sustavnim pregledima, kako Cochraneovih, tako i ne-Cochraneovih. RoB iz 2011 koristi se i u većini ne-Cochraneovih sustavnih pregleda (85). Temeljem naših podataka, nakon što se alat RoB 2 počne koristiti u svim Cochraneovim pregledima, moći ćemo u budućnosti usporediti je

li se s novim alatom poboljšala prikladnost prosudbi Cochraneovih autora za ovu domenu. Stoga naši rezultati mogu biti korisni autorima koji će i dalje koristiti alat RoB iz 2011., kao i iz perspektive metodologije istraživanja za mjerenje adekvatnosti metoda u novom alatu RoB 2 i usporedbu s rezultatima za stari alat. Nove metode istraživanja nastavljaju se razvijati, ali važno je osigurati da su bolje od prethodnih metoda.

Naše je istraživanje još jedna potvrda da više pozornosti treba posvetiti metodama koje koriste Cochraneovi autori. Pogrešne procjene RoB-a dovest će do pogrešnih zaključaka, što u konačnici može poslati pogrešne poruke korisnicima - zdravstvenim radnicima i donositeljima odluka. Urednici i recenzenti mogli bi pomoći u osiguravanju odgovarajuće upotrebe sustavnih metoda pregleda.

3.2. Ograničenja

U ovom smo istraživanju koristili računalni program za izdvajanje podataka, ali moguće je da smo pogriješili u interpretaciji podataka. Pokušali smo izbjeći ovu mogućnost dvostrukom provjerom svih odluka dvojice neovisnih autora. Također trebamo navesti da u našem istraživanju nismo pokušali otkriti moguću pristranost u izvornim istraživanjima, na primjer pomoću Berger-Exnerovog testa (109-111), jer nismo imali odgovarajući ulaz za izvršenje. To je bio slučaj čak i u našim prethodnim radovima o pristranosti koja proizlazi iz neadekvatnog generiranja nasumičnog redoslijeda (83) i prikrivanja razvrstavanja (84). Naš skup komentara i procjena autora CSR, na kojima smo izvršili analizu, nije obuhvaćao analizu cjelovitih tekstova izvornih istraživanja. Umjesto toga, oslanjali smo se samo na sadržaj RoB tablica iz CSR-ova, a ove tablice uključuju samo informacije koje su autori CSR u njima izvijestili. To bi moglo pomoći u provjeri je li procjena autora CSR zapravo bila primjerena, kad autori nisu pružili odgovarajuće objašnjenje u popratnom komentaru.

4. Prikladnost procjena Cochraneovih autora u istraživanju o kirurškim intervencijama

Glavni rezultat ovog istraživanja je da su RoB procjene za randomizaciju, prikrivanje razvrstavanja i domene zasljepljivanja točnije u Cochraneovim pregledima koji su procjenjivali kirurška istraživanja u usporedbi s pregledima ne-kirurških istraživanja. Iako su sedam domena obvezni dijelovi Cochraneovog RoB alata za 2011. godinu, neke od četiri analizirane domene često su nedostajale u analiziranim pregledima literature. Odsutnost analiziranih četiriju domena bila je češća u pregledima ne-kirurških istraživanja.

U ovim istraživanjima odlučili smo analizirati samo Cochraneove sustavne preglede iz dva razloga. Prvo, Cochraneovi pregledi moraju slijediti Cochraneove metode, a upotreba Cochrane RoB alata je za njih obavezna. Drugo, prethodno smo pokazali da se većina autora ne-Cochraneovih sustavnih pregleda koristila procjenom RoB-a, a većina tih su koristili Cochraneov RoB alat; međutim, većina ih je koristila neadekvatno (85). U ranijem istraživanju naše istraživačke skupine, među 269 analiziranih ne-Cochraneovih sustavnih pregleda koji su koristili Cochraneov RoB alat, samo 16 (5,9%) je u potpunosti opisalo RoB rezultate, tj. izvijestilo je i o procjeni i o popratnom komentaru koji podupire procjenu (85). Zbog neadekvatnog prikazivanja podataka, odnosno izostanka prikazivanja i procjene i komentara s obrazloženjem, analiza prikladnosti procjene RoB-a otežana je u sustavnim pregledima koji nisu Cochraneovi. Unatoč tome, naša su otkrića relevantna za Cochraneove i ne-Cochraneove sustavne preglede, upravo zato što većina sustavnih pregleda koji nisu Cochraneovi koriste Cochraneov RoB alat.

Odsutnost nekih RoB domena u analiziranim Cochrane pregledima ukazuje na to da su autori Cochranea odlučili "prilagoditi" Cochraneov RoB alat uklanjanjem nekih domena iz zadanih postavki RoB tablice. Ovo nije jedina prilagodba koju smo primijetili. Također smo otkrili da su mnogi autori Cochranea uveli poddomene, tj. više procjena za jednu domenu, na temelju različitih ishoda. U tim su slučajevima za jednu RoB domenu autori donijeli više procjena na temelju vrsta ishoda, na primjer - jednu procjenu za objektivne ishode i jednu za subjektivne ishode. Sljedeći je primjer prilagodbe kada autori podijele domenu za zasljepljivanje ispitanika i osoblja na dvije domene - jednu za zasljepljivanje ispitanika i drugu za zasljepljivanje osoblja. Obrazloženje ove prilagodbe su različiti ishodi, a različiti ključni pojedinci koji sudjeluju u istraživanju mogu dati različite ocjene RoB-a.

Stoga smo pronašli veći broj procjena u odnosu na broj analiziranih RoB tablica, tj. broj pojedinačnih istraživanja analiziranih s tim RoB domenama. Naročito je veći broj procjena

pronađen u obje RoB domene za zasljepljivanje, što ukazuje da su autori Cochranea željeli naglasiti potencijalnu razliku u utjecaju uspjeha zasljepljivanja na različite vrste ishoda.

Otkrili smo da je RoB domena za zasljepljivanje ispitanika i osoblja imala najmanju učestalost prikladnih procjena. Nadalje, otkrili smo da obje domene o zasljepljivanju imaju veću učestalost prikladnih procjena u kirurškim istraživanjima. Za domenu zasljepljivanja ispitanika i osoblja to bi moglo biti zbog toga što je zasljepljivanje tih pojedinaca teško postići u kirurškim pokusima (92, 112), što dovodi do veće transparentnosti u opisima metodologije u kirurškim pokusima. Stoga se automatski, u kirurškim pokusima, manje procjenjuje "nizak rizik" pristranosti, koji je povezan s najmanjom učestalošću prikladnih procjena.

Rezultati naših prethodnih istraživanja o procjenama RoB-a u sustavnim pregledima pokazali su da bi bilo korisno podijeliti domenu „zasljepljivanje ispitanika i osoblja“ na dvije domene, jednu za ispitanike i jednu za osoblje (82). To je ugrađeno u alatu RoB 2 (78), koji još nije implementiran u sve Cochrane protokole i preglede. Također smo otkrili da ista ista domena neće imati koristi od daljnjeg razdvajanja domene na temelju različitih ishoda (82).

Suprotno tome, za domenu u pogledu zasljepljivanja osoba koje bilježe ishode otkrili smo da bi bilo korisno uvesti poddomene za objektivne naspram subjektivnih ishoda. Ovaj bi pristup smanjio broj nedefiniranih ishoda s naknadnim povećanjem učestalosti prikladnih procjena (87). Također smo utvrdili da duljina komentara utječe na ispravno opravdanje procjene RoB-a i njezinu primjerenost (82).

Naši rezultati daju dva aspekta preporuka, uključujući preporuke za provođenje pokusa s kirurškim intervencijama (kako bi se smanjio rizik od pristranosti) i praktična rješenja za alate za procjenu RoB-a (kako bi se osigurale odgovarajuće procjene RoB-a o istraživanjima).

Iako ovo istraživanje nije imalo za cilj analizirati metodološke nedostatke kirurških istraživanja, postoje neke jednostavne preporuke koje se mogu generalizirati. Redoslijed raspodjele ispitanika uvijek treba biti nasumičan. Preporučuje se računalna randomizacija jer ima višestruke prednosti. Ako se uzme u obzir randomizacija u blokovima, blokovi bi trebali biti veći (izbjegavati četvorke), savjetuje se korištenje minimizacije za višestruku stratifikaciju, a svaku podskupinu treba randomizirati odvojeno. Još jedna prednost računalne randomizacije je njena dostupnost. Središnja randomizacija od treće strane također doprinosi prikrivanju razvrstavanja koje je osigurano do kraja istraživanja. Stoga bi se uporaba neprozirnih zatvorenih omotnica s rednim brojevima trebala napustiti [10]. Ako autori sustavnih pregleda u radovima ne pronađu informacije o određenim metodama za randomizaciju i prikrivanje razvrstavanja, trebali bi biti oprezni kako bi izbjegli pogrešnu procjenu RoB-a.

Ako nije moguće zaslijepiti ključne osobe, kao što je to slučaj u mnogim istraživanjima kirurških intervencija, treba poduzeti korake kako bi se smanjio rizik od pristranosti na različitim razinama. Na razini ispitanika, ako je lažna (engl. *sham*) operacija neetična ili nije odobrena, a ograničavanje informacija o postupku na pacijenta nije potrebno, prilikom planiranja istraživanja trebaju se primijeniti neke jednostavne mjere, kao i detaljan opis koji podupire procjenu RoB (113). Te mjere mogu uključivati nemiješanje skupina pacijenata, prikrivanje rezova većim oblogama i pružanje definiranog standarda skrbi jednakog i za intervencijske i za kontrolne skupine. Posljednje dvije pomažu smanjiti RoB kad zasljepljivanje kirurga nije moguće. Međutim, postavljanje temeljeno na ekspertizi može se koristiti s više kirurga / timova koji izvode isti postupak za istu grupu. Sve se to može dodati u računalni program za analizu RoB-a kako bi se poboljšala konačna procjena o RoB-u za ove domene.

Za procjenu ishoda, dostupnost zaslijepljenog sekundarnog tima kirurga ili kirurških sestara presudna je za postizanje niskog RoB-a. Međutim, u neslijepoj procjeni, osjetljivost ishoda na nedostatak zasljepljivanja najvažniji je čimbenik za konačnu RoB procjenu. Stoga bi svi ishodi trebali biti definirani (opis pozitivnog i negativnog događaja / kriterija) prije početka istraživanja. Kada definirani ishod nije objektivan, osoba koja bilježi ishode treba unaprijed odrediti i način mjerenja ishoda adresirati prije bilježenja opažanja (92). Ako ništa od navedenog nije moguće, treba ponoviti procjenu ako je poželjno ili barem izjavu koja priznaje ograničenja. Potencijalno, uvođenje padajućeg izbornika s različitim vrstama ishoda u računalni program za provođenje sustavnih pregleda moglo bi pomoći u boljoj procjeni RoB-a u istraživanjima s različitim skupinama ishoda.

Praktična rješenja koja ovo istraživanje može ponuditi za poboljšanje RoB prosudbi u Cochraneovim sustavnim pregledima uključuju prijedlog da prilagođavanje RoB tablice ne smije biti dopušteno u programu RevMan koji se koristi za izradu Cochraneovih sustavnih pregleda. Inače, autori će i dalje imati mogućnost brisanja određenih RoB domena koje možda smatraju nebitnima. Potpuno izvješćivanje o RoB alatu podrazumijeva upotrebu svih RoB domena, a procjena i komentari za svaku domenu važni su za prikladnu procjenu metodologije istraživanja. Nadalje, naši rezultati u vezi s duljinom komentara ukazuju na to da autore treba poticati i upozoravati pomoću računalnog programa da pruže detaljnije opise svojih procjena u polju za komentare u tablici RoB. Intervencije za poboljšanje ocjene urednika i recenzenata procjena RoB-a, kao i intervencije za poboljšanje ocjene autora, bile bi dobrodošle. Procjene RoB-a koriste se za davanje zaključaka o pregledu i u GRADE pristupu za ocjenu sigurnosti dokaza u sustavnim pregledima. Dakle, neprikladne procjene RoB-a mogu se pretvoriti u

neprikladne zaključke pregleda i neprikladnu procjenu sigurnosti dokaza, što će dovesti do pogrešnih preporuka za daljnja istraživanja i praksu.

Ograničenje našeg istraživanja je da smo možda napravili nenamjerne pogreške kada smo procjenjivali primjerenost procjena Cochraneovih autora analizom objavljenih popratnih komentara. Kako bismo smanjili pristranost, izvršili smo neovisnu procjenu dva autora za svaku analiziranu domenu i poddomenu. Uz to, u analizu smo uvrstili samo prve četiri domene Cochrane RoB alata, jer su upute iz Cochrane priručnika za ove četiri domene bolje okarakterizirane u odnosu na preostale tri domene (79, 80, 86).

Nadalje, primarni cilj istraživanja bio je procijeniti razlike u broju i primjerenosti procjena RoB-a u istraživanjima s kirurškom intervencijom. Stoga je kategorizacija intervencija u skupinu "kirurških nasuprot ne-kirurških" izabrana prema Kappa statistici kao mjera međusobnog dogovora. Iako se nismo usmjerili na stvarnu razinu dogovora, koristili smo je kao mjeru za bolju definiciju skupina koje se uspoređuju. Međutim, moramo naglasiti da je glavni nedostatak ove metode činjenica da ona može dovesti do onog što se naziva "problemom osnovne stope", a osjetljiva je na "istinsku prevalenciju" u podatcima. Ako je istinska prevalencija populacije visoka ili niska, slučajno očekivani sporazum se povećava, a veličina Kappe opada. Štoviše, unutar široke kategorije ne-kirurških intervencija postoji mnogo intervencija (npr. psihosocijalne intervencije, psihoterapije, probir itd.) koje se ne mogu zaslijepiti pa dijele iste probleme za procjenu RoB-a s kirurškim intervencijama. Stoga bi bilo dobro nastaviti daljnja istraživanja poteškoća povezanih s procjenom RoB-a u drugim intervencijama koje bi moglo biti teško zaslijepiti.

ZAKLJUČAK

Raspodjela različitog rizika od pristranosti po kategorijama (visok / nizak / nejasan rizik) koje su Cochraneovi autori dodijelili za kirurška i ne-kirurška istraživanja nije se razlikovala za domene generiranja nasumičnog redoslijeda i raspodjele ispitanika, ali su se značajno razlikovale u dvije domene vezane za zasljepljivanje ključnih pojedinaca.

Raspodjela novih procjena rizika od pristranosti koje su napravljene u okviru ove disertacije, na temelju objašnjenja iz RoB tablica, značajno se razlikovala između kirurških i ne-kirurških istraživanja u domeni za generiranje nasumičnog redoslijeda i domeni za zasljepljivanje osoba koje bilježe ishode.

Učestalost prikladnih procjena rizika od pristranosti značajno je varirala između prve četiri RoB domene. U cijelom uzorku analiziranih pregleda, najveća učestalost prikladnih procjena Cochraneovih autora pronađena je u RoB domeni za generiranje nasumičnog redoslijeda (87,9%), nakon čega slijedi domena za zasljepljivanje osoba koje bilježe ishode (72,9%), prikrivanje razvrstavanja (71,9%) i zasljepljivanje ispitanika i osoblja (70,7%).

Učestalost prikladnih procjena RoB-a za sve analizirane RoB domene uglavnom je bila veća u kirurškim pokusima nego u ne-kirurškim pokusima. Za dvije RoB domene koje procjenjuju zasljepljivanje o domenu za raspodjelu ispitanika, ova razlika između kirurških i ne-kirurških istraživanja bila je statistički značajna, dok razlika između dvije vrste istraživanja nije bila značajna za RoB domenu koja procjenjuje prikladnost generiranja nasumičnog redoslijeda.

Raspodjela različitog rizika od kategorija pristranosti (visokog / niskog / nejasnog) koje su autori CSR dodijelili za kirurška i ne-kirurška istraživanja nije se razlikovala za prve dvije domene (generiranje nasumičnog redoslijeda i prikrivanje razvrstavanja), ali su se značajno razlikovale za dvije domene koje procjenjuju zasljepljivanje ključnih skupina osoba.

U analiziranim Cochraneovim sustavnim pregledima učestalost neprikladnih procjena u analizirane prve četiri domene RoB alata iz 2011. bila je vrlo visoka. Postotak procjena Cochraneovih autora koje nisu bile u skladu s Cochraneovim priručnikom za izradu sustavnih pregleda u pokusima uključenim u analizirane Cochraneove sustavne preglede bio je 12% za domenu o generiranju nasumičnog redoslijeda, 29% za domenu o razvrstavanju ispitanika, 26% za domenu zasljepljivanja ispitanika i osoblja te 22% za domenu o zasljepljivanju osoba koje bilježe ishode.

KRATKI SAŽETAK

Uvod: Procjena valjanosti RCT-ova uključenih u Cochraneove sustavne preglede temelji se na alatu koji se naziva Cochraneov alat za procjenu rizika od pristranosti. Ranijim je istraživanjima uočeno da Cochraneovi autori često rade pogreške u procjeni rizika od pristranosti (engl. risk of bias, RoB) u objavljenim Cochraneovim sustavnim pregledima.

Cilj: Glavni cilj ove disertacije bio je analizirati učestalost različitih procjena RoB kroz prve četiri domene Cochraneovog RoB iz 2011. u RCT-ovima s intervencijama kirurškog tipa. Drugi glavni cilj je analiza podudarnosti procjena Cochraneovih autora s našim procjenama RoB u RCT-ovima s intervencijama kirurškog tipa kroz ove četiri odabrane domene koje su jasnije definirane u Cochraneovom priručniku. Sporedni cilj je analiza i usporedba učestalosti procjena RoB od strane Cochraneovih autora, a i naših procjena RoB između RCT-ova s kirurškim intervencijama i onih s ne-kirurškim intervencijama. Još jedan sporedni cilj je bilo otkrivanje mogućih generaliziranih pogreški u procesima procjene.

Metode: Analizirane su metode Cochraneovih sustavnih pregleda koji su uključili randomizirane kontrolirane pokuse, i koji su objavljeni 2015-2016. godine. Analizirani su podatci iz RoB tablica u Cochraneovim pregledima (RoB procjene i prateći komentari s objašnjenjima) za prve četiri RoB domene iz Cochraneova RoB alata iz 2011. godine: generiranje nasumičnog redoslijeda, prikrivanje razvrstavanja, zasljepljivanje ispitanika i osoblja, i zasljepljivanje osoba koje bilježe ishode. Odgovarajuće procjene definirane su kao one koje su u skladu s uputama iz Cochraneova priručnika za sustavne preglede intervencija. Učestalost prikladnih procjena uspoređivana je u kirurškim i ne-kirurškim istraživanjima te za po pojedinim domenama RoB procjene.

Rezultati: Usporedba RoB procjena za kliničke pokuse uključene u Cochraneove sustavne preglede koji su analizirali kirurške i ne-kirurške intervencije pokazala je kako se učestalost prikladnih procjena RoB-a koje su donosili autori Cochrane kretala od 87,9, 95% CI (87,3 do 88,6%) za randomizaciju do 70,7, 95% CI (69,8 do 71,5%) za zasljepljivanje ispitanika i osoblja. Za sve analizirane RoB domene, učestalost prikladnih procjena RoB domena bila je veća u kirurškim pokusima nego u ne-kirurškim pokusima. Za dvije RoB domene koje procjenjuju zasljepljivanje, razlika između kirurških i ne-kirurških istraživanja bila je statistički značajna ($P < 0{,}001$), dok razlika nije bila značajna za RoB domenu s obzirom na randomizaciju ($P = 0{,}124$) i prikrivanje razvrstavanja ($P = 0{,}039$, $\beta < 0{,}8$).

U analiziranim Cochraneovim sustavnim pregledima učestalost neprikladnih procjena u analizirane prve četiri RoB domene bila je vrlo visoka. Postotak procjena Cochraneovih autora koje nisu bile u skladu s Cochraneovim priručnikom za izradu sustavnih pregleda u pokusima uključenim u analizirane Cochraneove sustavne preglede bio je 12% za domenu o generiranju nasumičnog redoslijeda, 29% za domenu o razvrstavanju ispitanika, 26% za domenu zasljepljivanja ispitanika i osoblja te 22% za domenu o zasljepljivanju osoba koje bilježe ishode.

Zaključak: Procjene rizika od pristranosti bile su češće u skladu s uputama iz Cochraneova priručnika kada su Cochraneovi sustavni pregledi analizirali kirurške intervencije, u usporedbi s onima koji su analizirali ne-kirurške intervencije. Međutim, mnoge procjene RoB-a u Cochraneovim pregledima i kirurških i ne-kirurških istraživanja nisu u skladu s Cochraneovim priručnikom; stoga su potrebni daljnji koraci za detaljnu provjeru i poboljšanje procjene RoB-a u sustavnim pregledima o intervencijama.

TITLE AND ABSTRACT OF DOCTORAL THESIS

Title: Adequacy of risk of bias assessment in surgical vs non-surgical trials in Cochrane reviews

Background: Risk of bias (RoB) assessment is the crucial part of systematic review methodology. Bias in randomized controlled trials (RCTs) can lead to underestimating or overestimating interventions' actual effects. RCTs of surgical interventions may suffer from RoB that is avoidable in trials of other interventions, and vice versa.

Aim: This doctoral thesis aimed to compare RoB assessments' adequacy in surgical versus non-surgical RCTs included in published Cochrane reviews, to assess the most common differences in those RoB assessments, and to analyze the frequency of adequate RoB assessments in the first four domains of the Cochrane RoB tool.

Methods: RoB methods of published Cochrane systematic reviews were analyzed. Data were extracted from RoB tables in Cochrane reviews (judgments and accompanying explanatory comment) for the first four RoB domains used in the 2011 Cochrane RoB tool: randomization, allocation concealment, blinding of participants and personnel, and blinding of outcome assessors. We defined adequate assessments as those in line with instructions from the Cochrane Handbook for Systematic Reviews of Interventions. The frequency of adequate assessments was compared in surgical versus non-surgical trials. The most common differences in both groups of reviews were presented. The frequency of adequate assessments in each analyzed RoB domain was presented.

Results: In 729 analyzed Cochrane reviews, there were 10537 included RCTs. The frequency of adequate RoB judgments made by Cochrane authors ranged from 87,9% (95% CI 87,3 to 88,6%) for randomization to 70,7% (95% CI 69,8 to 71,5%) for blinding of participants and personnel. For the four analyzed RoB domains, the frequency of adequate RoB judgments was higher in surgical trials than in non-surgical trials. For two RoB domains assessing blinding, this difference between surgical and non-surgical trials was significant ($P < 0,001$), while the difference was not significant for the RoB domain regarding randomization ($P = 0.,124$) and allocation concealment ($P = 0,039$, $\beta < 0,8$).

In the analyzed Cochrane systematic reviews, the frequency of inadequate judgments in the analyzed first four domains of the 2011 RoB tool was very high. The percentage of Cochrane authors' judgments that were not in line with the Cochrane Handbook was 12% for

randomization, 29% for allocation concealment, 26% for blinding of participants and personnel, and 22% for blinding of outcome assessors.

Conclusions: The RoB judgments were more frequently in line with instructions from the Cochrane Handbook when Cochrane reviews assessed surgical trials, compared to those that analyzed non-surgical interventions. However, considering the high frequency of inadequate RoB judgments in published Cochrane reviews, further steps are warranted to scrutinize and improve RoB assessment in systematic reviews of interventions.

DODATNI MATERIJALI

Tablica 23 – Ishodi, hipoteze, korišteni statistički testovi i rezultati u radu o II. RoB domeni

Varijabla/Ishod	Nul-hipoteza	Metoda analize	Rezultat	Zaključak
Glavna analiza: broj i prikladnost procjena za rizik od pristranosti u domeni o zasljepljivanju ispitanika i osoblja	Procjene koje su napravili autori Cochraneovih sustavnih pregleda sukladne su našem izračunu prema Cochraneovom priručniku	Wilcoxonov test (upareni uzorci)	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina.
Sekundarna analiza: Duljina komentara i procjena	Duljina komentara ista je za sve vrste procjena u dodijeljenoj procjeni (visok, nizak ili nejasan rizik od pristranosti)	Kruskal-Wallisov test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina. Alternativna hipoteza prihvaćena. Duljina komentara utječe na dodijeljenu procjenu.
	Duljina komentara ista je za sve vrste procjena u izračunatoj procjeni (visok, nizak ili nejasan rizik od pristranosti)	Kruskal-Wallisov test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina. Alternativna hipoteza prihvaćena. Duljina komentara utječe na izračunatu procjenu.
	Duljina komentara je ista u grupi kada je postignuto zasljepljivanje, kao i u skupini kada zasljepljivanje nije postignuto	Kruskal-Wallisov test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina. Alternativna hipoteza prihvaćena. Duljina komentara utječe na postizanje zasljepljivanja.
	Duljina komentara je ista u podskupini koja se mijenja od dodijeljene nejasne procjene RoB u izračunatu visoku ocjenu RoB kao i u grupi s nepromijenjenom visokom ocjenom RoB	Mann-Whitneyev test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina. Alternativna hipoteza prihvaćena. Skraćenje komentara ne smanjuje rizik od pristranosti od visokog u nejasni.
	Duljina komentara je ista u podskupini koja se mijenja od dodijeljene procjene nejasnog RoB do izračunate procjene visokog RoB kao i u skupini s nepromijenjenom nejasnom procjenom RoB	Mann-Whitneyev test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina. Alternativna hipoteza prihvaćena. Skraćenje komentara ne smanjuje rizik od pristranosti od visokog u nejasni.
Sekundarna analiza: podijeljeni ishodi	Udio slučajeva s malo vjerojatnoćom nedostatka zasljepljujućeg utjecaja na ishod isti je u podskupini s podijeljenim ishodima kao u glavnom uzorku	Hi-kvadrat test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina. Alternativna hipoteza prihvaćena. Slučajevi imaju sklonost grupiranja u uzorku s podijeljenim ishodima.
Sekundarna analiza: podijeljena domena	Raspodjela dodijeljenih procjena jednaka je u podskupini (osoblju) kao u glavnom uzorku	Mann-Whitneyev test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina.

	Raspodjela izračunatih procjena jednaka je u podskupini (osoblju) kao u glavnom uzorku	Mann-Whitneyev test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina.
	Raspodjela dodijeljenih procjena jednaka je u podskupini (ispitanici) kao i u glavnom uzorku	Mann-Whitneyev test	p=0,0612	Nije nađena značajna razlika među skupinama.
	Raspodjela izračunatih procjena jednaka je u podskupini (ispitanici) kao i u glavnom uzorku	Mann-Whitneyev test	p=0,1412	Nije nađena značajna razlika među skupinama.
	Raspodjela dodijeljenih procjena jednaka je u spojenoj podskupini (ispitanici i osoblje) kao i u glavnom uzorku	Mann-Whitneyev test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina.
	Raspodjela izračunatih procjena jednaka je u spojenoj podskupini (ispitanici i osoblje) kao i u glavnom uzorku	Mann-Whitneyev test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina.
	Udio nejasnih procjena jednak je u podskupini (ispitanici) kao u glavnom uzorku	Hi-kvadrat test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina.
	Udio nejasnih procjena jednak je u podskupini (osoblje) kao u glavnom uzorku	Hi- kvadrat test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina.
	Udio nejasnih procjena jednak je u podskupini (ispitanici) kao i u podskupini (osoblje)	Hi- kvadrat test	p=0,1023	Statistički test ne pokazuje značajnu razliku među skupinama. Uzorak premalen.
	Procjene u podskupini ispitanika koje su dodijelili autori pregleda sukladne su našem izračunu prema Cochraneovom priručniku	Wilcoxonov test (upareni uzorci)	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina.
	Procjene u podskupini osoblja koje su dodijelili autori pregleda sukladne su našem izračunu prema Cochraneovom priručniku	Wilcoxonov test (upareni uzorci)	p=0,2565	Nije nađena značajna razlika među skupinama.
	Procjene za združenu domenu (ispitanici i osoblje) koje su dodijelili autori pregleda sukladne su našem izračunu prema Cochraneovom priručniku	Wilcoxonov test (upareni uzorci)	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina.
	Udio suglasnih procjena jednak je u podskupini (ispitanici) kao i u podskupini (osoblje)	Hi- kvadrat test	p<0,05	Statistički test pokazuje moguću razliku među skupinama. Uzorak premalen.
	Udio podudarnih procjena jednak je u spojenoj podskupini (ispitanici i osoblje) kao u glavnom uzorku	Hi- kvadrat test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina.

Tablica 24 – Ishodi, hipoteze, korišteni statistički testovi i rezultati u radu o III. RoB domeni

Varijabla/ishod	**Nulta hipoteza**	**Metoda analize**	**Rezultat**	**Zaključak**
Glavna analiza: Prikladnost procjena za rizik od pristranosti povezan sa zasljepljivanjem osoba koje bilježe ishode	Procjene Cochraneovih autora podudarne su s našim novim procjenama prema Cochraneovom priručniku	Wilcoxonov test (upareni uzorci)	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina.
Sekundarna analiza: raspodjela ishoda kad je domena podijeljena	Raspodjela procijenjenih ishoda za 6 kategorija ishoda ista je u pod-uzorku kad je domena podijeljena kao i u cijelom uzorku	Mann-Whitneyev test	p<0,05	Odbačena nulta hipoteza: značajna razlika između skupina.
Sekundarna analiza: prikladnost procjena prema kategoriji ishoda	Udio podudarnih procjena jednak je u podskupini ishoda (procjena kliničara) kao i u cijelom uzorku: 84,3% vs 77,9%	Hi- kvadrat test	p<0,05	Nulta hipoteza odbačena: značajna razlika među skupinama. Alternativna hipoteza prihvaćena: podskupina ishoda koje procjenjuje kliničar ima veći udio prikladnih procjena nego cijeli uzorak.
	Udio podudarnih procjena jednak je u podskupini ishoda (koje procjenjuje pacijent) kao i u cijelom uzorku: 84,4% vs 77,9%	Hi- kvadrat test	p<0,05	Nulta hipoteza odbačena: značajna razlika među skupinama. Alternativna hipoteza prihvaćena: podskupina ishoda koje procjenjuju pacijenti ima veći udio prikladnih procjena nego cijeli uzorak.
	Udio podudarnih procjena u podskupini ishoda (subjektivni ishodi) isti je kao u cijelom uzorku: 91,5% vs 77,9%	Hi- kvadrat test	p<0,05	Nulta hipoteza odbačena: značajna razlika među skupinama. Alternativna hipoteza prihvaćena: podskupina subjektivnih ishoda ima veći udio prikladnih procjena nego cijeli uzorak.
	Udio podudarnih procjena isti je u podskupini ishoda (objektivni ishodi) kao u cijelom uzorku: 71,5% vs 77,9%	Hi- kvadrat test	p<0,05	Nulta hipoteza odbačena: značajna razlika među skupinama. Alternativna hipoteza prihvaćena: podskupina objektivnih ishoda ima niži udio prikladnih procjena u usporedbi s cijelim uzorkom.
	Udio podudarnih procjena isti je u pod-uzorku procjena kad je domena podijeljena kao u cijelom uzorku: 80,6% vs 77,9%	Hi- kvadrat test	p=0,013 5	Statistički test pokazuje moguće razlike između skupina; uzorak je premalen (β>0,2).
Sekundarna analiza. Svi oblici subjektivnih kategorija ishoda	Udio podudarnih procjena isti je u spojenim podskupinama subjektivnih ishoda (procjena kliničara, procjena pacijenta,	Hi- kvadrat test	p<0,05	Nulta hipoteza odbačena: značajna razlika među skupinama. Alternativna hipoteza prihvaćena: spojena podskupina subjektivnih ishoda (procjena kliničara,

	subjektivni) kao u cijelom uzorku: 86,9% vs 77,9%			procjena pacijenata i subjektivni) ima veći udio prikladnih procjena u odnosu na cijeli uzorak.
	Udio podudarnih procjena isti je u spojenim podskupinama subjektivnih ishoda (procjena kliničara, procjena pacijenta, subjektivni) kad je domena podijeljena, kao u pod-uzorku kad je domena podijeljena: 85,7% vs 80,6%	Hi- kvadrat test	p<0,05	Nulta hipoteza odbačena: značajna razlika među skupinama. Alternativna hipoteza prihvaćena: spojena podskupina subjektivnih ishoda (procjena kliničara, procjena pacijenata i subjektivni) kad je domena podijeljena ima veći udio prikladnih procjena nego pod-uzorak kad je domena podijeljena.
	Udio podudarnih procjena isti je u spojenoj podskupini subjektivnih ishoda (procjena kliničara, procjena pacijenta, subjektivni) kad je domena podijeljena i procjene su iste kao u pod-uzorku kad je domena podijeljena i procjene su sve jednake: 90,3% vs 92,2%	Hi- kvadrat test	p=0,3907	Prihvaćena nulta hipoteza: nema značajnih razlika između skupina.
	Udio podudarnih procjena jednak je u spojenoj podskupini subjektivnih ishoda (procjena kliničara, procjena pacijenta i subjektivni) kada je domena podijeljena i procjene se razlikuju kao u cijelom pod-uzorku kada je domena podijeljena i procjene se razlikuju: 83,4% vs 75,4%	Hi- kvadrat test	p<0,05	Nulta hipoteza odbačena: značajna razlika među skupinama. Prihvaćena alternativna hipoteza: spojena podskupina subjektivnih ishoda (procjena kliničara, procjena pacijenta i subjektivni) kada je domena podijeljena i procjene se razlikuju, ima veći udio prikladnih procjena od cijelog poduzorka kada je domena podijeljena i procjene se razlikuju.
Sekundarna analiza: iste procjene za sve podijeljene ishode	Udio podudarnih procjena jednak je u poduzorku procjena kada je domena podijeljena i procjene su iste kao u cijelom uzorku	Hi- kvadrat test	p<0,05	Nulta hipoteza odbačena: značajna razlika među skupinama. Prihvaćena alternativna hipoteza: poduzorak procjena kada je domena podijeljena i procjene su iste ima veći udio prikladnih procjena od cijelog uzorka.
	Udio podudarnih procjena jednak je u poduzorku procjena kada je domena podijeljena i procjene se razlikuju kao u cijelom uzorku	Hi- kvadrat test	p=0,0553	Statistički test ne pokazuje značajnu razliku među skupinama.

Tablica 25 – Ishodi, hipoteze, korišteni statistički testovi i rezultati u radu o istraživanjima s kirurškim i ne-kirurškim intervencijama

Varijabla/ishod	Nulta hipoteza	Metoda analize	Rezultat	Zaključak
Glavna analiza: Ukupni udio odgovarajućih RoB procjena dodijeljenih od strane Cochrane autora	Udio prikladnih procjena RoB-a jednak je u skupinama pokusa s kirurškim (89,7%) naspram ne-kirurškim (87,8%) intervencijama u domeni za generiranje nasumičnog redoslijeda	Hi- kvadrat test	$P = 0,13$	Prihvaćena nulta hipoteza: nema značajne razlike među skupinama.
	Udio prikladnih procjena RoB-a jednak je u skupinama pokusa s kirurškim (75,2%) naspram ne-kirurškim (71,7%) intervencijama u domeni za prikrivanje razvrstavanja	Hi- kvadrat test	$P = 0,04$ $\beta < 0,8$	Test je neuvjerljiv: čini se da postoje prikladnije procjene RoB-a u kirurškoj nego u ne-kirurškoj skupini, međutim snaga testa je manja od 80%.
	Udio prikladnih procjena RoB-a jednak je u skupinama pokusa s kirurškim (79,5%) naspram ne-kirurškim (69,9%) intervencijama u domeni za zasljepljivanje ispitanika i osoblja i za zajedničku domenu	Hi- kvadrat test	$P < 0,001$	Odbačena nulta hipoteza: Postoje znatno prikladnije procjene RoB-a u kirurškoj nego u ne-kirurškoj skupini.
	Udio prikladnih procjena RoB-a jednak je u skupinama pokusa s kirurškim (81,1%) u odnosu na ne-kirurške (72,2%) zahvate u domeni za zasljepljivanje osoba koje bilježe ishode i zajedničku domenu zasljepljivanja	Hi- kvadrat test	$P < 0,001$	Odbačena nulta hipoteza: Postoje znatno prikladnije procjene RoB-a u kirurškoj nego u ne-kirurškoj skupini.
Sekundarna analiza: Broj i udio Cochraneovih sustavnih pregleda	Raspodjela kirurških vs. ne-kirurških pregleda unutar domene jednaka je između sve četiri RoB domene	Kruskal-Wallisov test	$P < 0,001$	Nulta hipoteza odbačena: značajna razlika među skupinama. Post-hoc Conoverov test nije otkrio razlike u parovima pa je ponovno primijenjena jednosmjerna ANOVA
	Raspodjela kirurških vs. ne-kirurških pregleda unutar domene jednaka je između sve četiri RoB domene.	Jednosmjerna ANOVA	$P < 0,001$	Nulta hipoteza odbačena: značajna razlika među skupinama. Post-hoc Student-Newman-Keulsov test otkriva da je domena za zasljepljivanje ispitanika i osoblja pokazala značajno različitu raspodjelu kirurških i ne-kirurških pregleda. Ova je domena bila prisutna u

				znatno manje sustavnih pregleda od ostalih domena
	Učestalost korištenja četiri analizirane RoB domene jednaka je između kirurških i ne-kirurških sustavnih pregleda.	Kruskal-Wallisov test	$P = 0{,}84$	Prihvaćena nulta hipoteza: nema značajne razlike među skupinama.
Sekundarna analiza: Broj i udio pokusa kojima nedostaju podaci za određene domene i istraživanja promatrana prema vrstama intervencija	Raspodjela kirurških vs. ne-kirurških istraživanja prema domenama koje nedostaju ili su uočene jednaka je između sve četiri RoB domene.	Kruskal-Wallisov test	$P < 0{,}001$	Nulta hipoteza odbačena: značajna razlika među skupinama. Post-hoc Conoverov test pokazuje da je domena za zasljepljivanje ispitanika i osoblja pokazala značajno različitu raspodjelu kirurških vs ne-kirurških pregleda. Ova je domena bila prisutna u znatno manje pokusa od drugih domena.
	Učestalost korištenja četiri analizirane RoB domene jednaka je između kirurških i ne-kirurških pokusa.	Kruskal-Wallisov test	$P < 0{,}001$	Nulta hipoteza odbačena: značajna razlika među skupinama. Post-hoc Conoverov test pokazuje da se skupina ne-kirurških istraživanja koja nema domenu razlikuje od svih ostalih skupina. Ova je skupina pokazala najviše varijabilnosti.
Sekundarna analiza: Broj i udio procjena analiziranih prema vrstama intervencija	Raspodjela procjena za procijenjena kirurška vs. ne-kirurška istraživanja jednaka je između sve četiri RoB domene.	Kruskal-Wallisov test	$P = 0{,}13$	Prihvaćena nulta hipoteza: nema značajne razlike među skupinama.
Sekundarna analiza: Raspodjela procjena o riziku (visok / nizak / nejasan) koje su dodijelili autori Cochranea	Raspodjela je ista u skupini istraživanja s kirurškim ili ne-kirurškim zahvatima u domeni za generiranje nasumičnog redoslijeda	Mann-Whitneyev test	$P = 0{,}41$	Prihvaćena nulta hipoteza: nema značajne razlike među skupinama.
	Raspodjela je ista u skupini istraživanja s kirurškim ili ne-kirurškim zahvatima u domeni za prikrivanje razvrstavanje	Mann-Whitneyev test	$P = 0{,}96$	Prihvaćena nulta hipoteza: nema značajne razlike među skupinama.
	Raspodjela je ista u skupini istraživanja s kirurškim ili ne-kirurškim zahvatima u domeni za zasljepljivanje ispitanika i osoblja	Mann-Whitneyev test	$P < 0{,}001$	Odbačena nulta hipoteza: U kirurškim pokusima manje je procjena o niskom RoB-u i više je procjena nejasnog rizika u odnosu na ne-kirurške pokuse
	Raspodjela je ista u skupini istraživanja s kirurškim ili ne-kirurškim zahvatima u domeni za zasljepljivanje	Mann-Whitneyev test	$P < 0{,}001$	Odbačena nulta hipoteza: manje je procjena niskog RoB-a i više procjena

	osoba koje bilježe ishode i zajedničku domenu za zasljepljivanje			nejasnog RoB-a u kirurškoj skupini nego u ne-kirurškoj
Sekundarna analiza: Raspodjela procjena o riziku (visok / nizak / nejasan) izračunata u našim istraživanjima	Raspodjela je ista u skupini istraživanja s kirurškim vs. ne-kirurškim zahvatima u domeni za generiranje nasumičnog redoslijeda	Mann-Whitneyev test	$P = 0{,}02$	Odbačena nulta hipoteza: Postoji više procjena o niskom i visokom RoB-u zbog manjeg broja procjena nejasnog RoB-u u kirurškoj vs. ne-kirurškoj skupini.
	Raspodjela je ista u skupini istraživanja s kirurškim vs. ne-kirurškim zahvatima u domeni za prikrivanje razvrstavanja	Mann-Whitneyev test	$P = 0{,}07$	Prihvaćena nulta hipoteza: nema značajne razlike među skupinama.
	Raspodjela je ista u skupini istraživanja s kirurškim vs. ne-kirurškim zahvatima u domeni za zasljepljivanje ispitanika i osoblja, i zajedničkoj domeni za zasljepljivanje	Mann-Whitneyev test	$P = 0{,}56$	Prihvaćena nulta hipoteza: nema značajne razlike među skupinama.
	Raspodjela je ista u skupini istraživanja s kirurškim vs. ne-kirurškim zahvatima u domeni za zasljepljivanje osoba koje bilježe ishode i zajedničkoj domeni za zasljepljivanje	Mann-Whitneyev test	$P < 0{,}001$	Odbačena nulta hipoteza: U kirurškoj vs. ne-kirurškoj skupini manje je procjena niskog RoB-a, a više je procjena nejasnog RoB-a
Sekundarna analiza: Uspješno zasljepljivanje ključnih pojedinaca	Učestalost uspješnog zasljepljivanja ispitanika i osoblja ista je u skupinama istraživanja s kirurškom intervencijom (4,0%) u odnosu na ne-kirurišku (12,1%) intervenciju.	Hi-kvadrat test	$P < 0{,}001$	Odbačena nulta hipoteza: Uspješno zasljepljivanje ispitanika i osoblja znatno je rjeđe u kirurškim vs. ne-kirurškim istraživanjima.
	Učestalost uspješnog zasljepljivanja ocjenjivača ishoda jednaka je u skupinama istraživanja s kirurškom intervencijom (9,1%) u odnosu na ne-kiruršku (15,7%) intervenciju.	Hi- kvadrat test	$P < 0{,}001$	Odbačena nulta hipoteza: Uspješno zasljepljivanje osoba koje bilježe ishode značajno je rjeđe u kirurškim vs. ne-kirurškim istraživanjima.

Tablica 26 – Raspodjela osnova za procjenu RoB i njihova interpretacija

	Osnova za opravdanost procjene RoB		**Intervencija**		**Test diskriminacije**	**Test definicije**	**Post-hoc usporedba parova**
Domena I (generiranje nasumičnog redoslijeda)	**Faktor broj**	**Kategorija popratnog komentara**	**Ne-kirurška**	**Kirurška**	**Mann-Whitney** ***P*** **< 0,001**	**ANOVA** ***P*** **= 0,001**	**Student-Newman-Keuls,** ***P*** **< 0,05 razlika od faktora**
	#1	Randomizacija nije opisana	5317 (57,0%)	389 (50,0%)	Razlika u raspodjeli kategorija između grupa	Razlika u udjelima između kategorija	#3 #5
	#2	Tablica nasumičnih brojeva	815 (8,7%)	68 (8,7%)			
	#3	Računalna randomizacija	2597 (27,8%)	253 (32,5%)			#1
	#4	Mehanička randomizacija	326 (3,5%)	33 (4,2%)			
	#5	Neodgovarajuća randomizacija	270 (2,9%)	35 (4,5%)			#1
Domena II (prikrivanje razvrstavanja)	**Faktor broj**	**Tip prikrivanja**	**Ne-kirurška**	**Kirurška**	**Mann-Whitney** ***P*** **= 0,48**	**Kruskal-Wallis** ***P*** **< 0,001**	**Conover** ***P*** **< 0,05**
	#1	Središnja raspodjela	506 (5,3%)	32 (4,1%)	Nema razlike u raspodjeli između grupa	Razlika u udjelu tipova prikrivanja razvrstavanja	#5
	#2	Nepotpun SNOSE	1117 (11,8%)	147 (18,6%)			#3
	#3	Nije jasno opisan	7200 (76,1%)	525 (66,5%)			#2 #5
	#4	Otvoreno razvrstavanje	386 (4,1%)	34 (4,3%)			#5
	#5	SNOSE	250 (2,6%)	51 (6,5%)			#1 #3 #4
Domena III+* (zasljepljivanje ispitanika i osoblja)	**Faktor broj**	**Zasljepljivanje**	**Ne-kirurška**	**Kirurška**	**Mann-Whitney** ***P*** **< 0,001**	**Kruskal-Wallis** ***P*** **< 0,001**	**Conover** ***P*** **< 0,05**
	#1	Postignuto	1239 (12,1%)	38 (4,0%)	Razlika u raspodjeli između grupa	Razlike u udjelima postizanja	#2 #3
	#2	Nije postignuto ili nije moguće	4318 (42,1%)	414 (43,8%)			#1
	#3	Vjerojatno postignuto	4709 (45,9%)	493 (52,2%)			#1
	Faktor broj	**Izostanak zasljepljivanja na ishod**	**Ne-kirurška**	**Kirurška**	**Mann-Whitney** ***P*** **< 0,001**	**Kruskal-Wallis** ***P*** **< 0,001**	**Conover** ***P*** **< 0,05**
	#1	Ima utjecaja	472 (4,6%)	21 (2,2%)	Razlika u raspodjeli između grupa	Razlike u udjelima utjecajnosti	#3
	#2	Nejasno ima li utjecaja	9340 (91,0%)	832 (88,0%)			#3
	#3	Nema utjecaja	454 (4,4%)	92 (9,7%)			#1 #2
Domena IV+* (zasljepljivanje osoba koje bilježe ishode)	**Faktor broj**	**Zasljepljivanje**	**Ne-kirurška**	**Kirurška**	**Mann-Whitney** ***P*** **< 0.001**	**Kruskal-Wallis,** ***P*** **< 0.001**	**Conover** ***P*** **< 0.05**
	#1	Postignuto	1722 (15,7%)	84 (9,1%)	Razlika u raspodjeli između grupa	Razlike u udjelima postizanja	#2 #3
	#2	Nije postignuto ili nije moguće	6695 (61,1%)	609 (65,9%)			#1
	#3	Vjerojatno postignuto	2540 (23,2%)	231 (25,0%)			#1
	Faktor broj	**Izostanak zasljepljivanja na ishod**	**Ne-kirurška**	**Kirurška**	**Mann-Whitney** ***P*** **= 0,21**	**Kruskal-Wallis,** ***P*** **< 0,001**	**Conover** ***P*** **< 0,05**
	#1	Ima utjecaja	919 (8,4%)	31 (3,4%)	Nema razlike u raspodjeli između grupa	Razlike u udjelima utjecajnosti	#2
	#2	Nejasno ima li utjecaja	9045 (82,5%)	839 (90,8%)			#1
	#3	Nema utjecaja	993 (9,1%)	54 (5,8%)			

** također uključuje podatke za zajedničke dvije domene za zasljepljivanje; SNOSE = uzastopno numerirane neprozirne zatvorene omotnice (engl. sequentially numbered sealed opaque envelopes)*

POPIS LITERATURE

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996;312(7023):71-2.
2. Medicine JH. Evidence-Based Medicine 2021 [Available from: https://www.hopkinsmedicine.org/gim/research/method/ebm.html.
3. Guyatt G. Evidence-based medicine. AC Journal Club. 1991;114(2).
4. White KL. Healing the Schism: Epidemiology, Medicine, and the Public's Health: New York (NY): Springer-Verlag; 1991.
5. Sur RL, Dahm P. History of evidence-based medicine. Indian J Urol. 2011;27(4):487-9.
6. Feinstein AR, Di Massa R. Prognostic significance of valvular involvement in acute rheumatic fever. The New England journal of medicine. 1959;260(20):1001-7.
7. Feinstein AR. Clinical epidemiology. I. The populational experiments of nature and of man in human illness. Annals of internal medicine. 1968;69(4):807-20.
8. Sackett DL. Clinical epidemiology. American journal of epidemiology. 1969;89(2):125-8.
9. How to read clinical journals: I. why to read them and how to start reading them critically. Canadian Medical Association journal. 1981;124(5):555-8.
10. Evidence-Based Medicine Working Group: Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA. 1992;268(17):2420-5.
11. Guyatt GHR, D. Users' Guides to the Medical Literature: A Manual of Evidence-Based Clinical Practice (third ed.). New York: McGraw-Hill.; 2015.
12. Rashid A, Finnikin S, Tackett S. Accreditation drives teaching: evidence-based medicine and medical education standards. BMJ evidence-based medicine. 2020.
13. Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine". The American journal of medicine. 1997;103(6):529-35.
14. Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine. Bmj. 1995;310(6987):1085-6.
15. Sackett DL. Surveys of self-reported reading times of consultants in Oxford, Birmingham, Milton-Keynes,

Bristol, Leicester, and Glasgow. Evidence-based medicine: London: Churchill Livingstone.

16. Cohen AM, Stavri PZ, Hersh WR. A categorization and analysis of the criticisms of Evidence-Based Medicine. International journal of medical informatics. 2004;73(1):35-43.

17. Szajewska H. Evidence-Based Medicine and Clinical Research: Both Are Needed, Neither Is Perfect. Annals of nutrition & metabolism. 2018;72 Suppl 3:13-23.
18. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plastic and reconstructive surgery. 2011;128(1):305-10.
19. Canadian Task Force on the Periodic Health Examination. The periodic health examination. . Canadian Medical Association journal. 1979;121(9):1193-254.
20. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest. 1989;95(2 Suppl):2s-4s.
21. American Society of Plastic Surgeons. Evidence Rating Scales. 2011 [Available from: https://www.plasticsurgery.org/documents/medical-professionals/health-policy/evidence-practice/ASPS-Rating-Scale-March-2011.pdf.
22. Centre for Evidence Based Medicine. [Available from: http://www.cebm.net.
23. American Society of Plastic Surgeons. Evidence-based clinical practice guidelines. 2011 [Available from: https://www.plasticsurgery.org/documents/Health-Policy/Guidelines/guideline-2011-reduction-mammaplasty.pdf.
24. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Bmj. 2008;336(7650):924-6.
25. ASPS: Evidence-Based Clinical Practice Guideline Methodology 2016 [Available from: https://www.plasticsurgery.org/documents/medical-professionals/quality-resources/ASPS-Evidence%E2%80%90Based-Clinical-Practice-Guideline-Methodology.pdf.
26. Armstrong R, Hall BJ, Doyle J, Waters E. Cochrane Update. 'Scoping the scope' of a cochrane review. Journal of public health (Oxford, England). 2011;33(1):147-50.
27. Bearman M, Dawson P. Qualitative synthesis and systematic review in health professions education. Medical Education. 2013;47(3):252-60.
28. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Information & Libraries Journal. 2009;26(2):91-108.
29. Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions. 2016.
30. Animated Storyboard: What Are Systematic Reviews? : Cochrane Consumers and Communication; [Available from: https://cccrg.cochrane.org/animated-storyboard-what-are-systematic-reviews.

31. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.
32. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, Brigham TJ. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. J Clin Epidemiol. 2015;68(6):617-26.
33. Higgins JPTT, J.; Chandler, J.; Cumpston, M.; Li, T.; Page, M. J.; Welch, V. A., editors. Cochrane Handbook for Systematic Reviews of Interventions. Chapter 4: Searching for and selecting studies. 2019 [Available from: https://training.cochrane.org/handbook/current/chapter-04#section-4-6.
34. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. PLoS Med. 2016;13(5):e1002028.
35. Boric K, Dosenovic S, Jelicic Kadic A, Batinic M, Cavar M, Urlic M, et al. Interventions for postoperative pain in children: An overview of systematic reviews. Paediatric anaesthesia. 2017;27(9):893-904.
36. Dosenovic S, Jelicic Kadic A, Miljanovic M, Biocic M, Boric K, Cavar M, et al. Interventions for Neuropathic Pain: An Overview of Systematic Reviews. Anesth Analg. 2017;125(2):643-52.
37. Boutron IP, M.J.; Higgins, J.P.T.; Altman, D.G.; Lundh, A; Hróbjartsson, A. Chapter 7: Considering bias and conflicts of interest among the included studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2020). Cochrane, 2020. Available from www.training.cochrane.org/handbook.
38. Cochrane [Available from: https://www.cochrane.org/.
39. Cochrane Library [Available from: https://www.cochranelibrary.com/.
40. Cochrane AL, Cox JG, Jarman TF. Pulmonary tuberculosis in the Rhondda Fach; an interim report of a survey of a mining community. British medical journal. 1952;2(4789):843-53.
41. Cochrane A. In Chalmers I, Enkin M, Keirse MJNC, eds. Effective care in pregnancy and childbirth. Oxford University Press, Oxford; 1989.
42. Roberts H. Daly, J, Evidence-based Medicine and the Search for a Science of Clinical Care. Sociology of Health & Illness. 2006;28(1):122-3.

43. Chalmers I, Enkin M, Keirse MJ. Effective care in pregnancy and childbirth: Pregnancy: Oxford University Press, USA; 1989.
44. Chalmers I, Hetherington J, Newdick M, Mutch L, Grant A, Enkin M, et al. The Oxford database of perinatal trials: Developing a register of published reports of controlled trials. Controlled Clinical Trials. 1986;7(4):306-24.
45. Ioannidis JPA. Meta-research: Why research on research matters. PLoS biology. 2018;16(3):e2005468.
46. Higgins JPT, Green S, (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration 2011. Available from http://handbook-5-1.cochrane.org/. Last accessed 14.01.2021.
47. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbé KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. J Clin Epidemiol. 1992;45(3):255-65.
48. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet. 2001;357(9263):1191-4.
49. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. Bmj. 1997;315(7109):617-9.
50. Sterne JA, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Statistics in medicine. 2002;21(11):1513-24.
51. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. Jama. 1995;273(5):408-12.
52. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
53. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials. 1995;16(1):62-73.
54. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17(1):1-12.
55. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ. 2001;323(7303):42-6.

56. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. Lancet. 2002;359(9305):515-9.
57. Schulz KF, Grimes DA. The Lancet handbook of essential concepts in clinical research. 2006.
58. Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. Lancet. 2002;359(9310):966-70.
59. Altman DG, Bland JM. How to randomise. Bmj. 1999;319(7211):703-4.
60. Higgins JPT, Green S, (editors). Assessing risk of bias in included studies - Part 2, Chapter 8 from Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration 2011. Available from https://https//handbook-5-1.cochrane.org/chapter_8/8_assessing_risk_of_bias_in_included_studies.htm. Last accessed 14.01.2021.
61. Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. Annals of internal medicine. 2002;136(3):254-9.
62. Schulz KF. Subverting Randomization in Controlled Trials. JAMA. 1995;274(18):1456-8.
63. Gøtzsche PC. Blinding during data analysis and writing of manuscripts. Control Clin Trials. 1996;17(4):285-90; discussion 90-3.
64. Haahr MT, Hróbjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. Clinical trials (London, England). 2006;3(4):360-5.
65. Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. BMJ (Clinical research ed). 2004;328(7437):432-.
66. Rees JR, Wade TJ, Levy DA, Colford JM, Jr., Hilton JF. Changes in beliefs identify unblinding in randomized controlled trials: a method to meet CONSORT guidelines. Contemporary clinical trials. 2005;26(1):25-37.
67. Hróbjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. International journal of epidemiology. 2007;36(3):654-63.
68. Sackett DL. Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't? International journal of epidemiology. 2007;36(3):664-5.
69. Pildal J, Hróbjartsson A, Jørgensen KJ, Hilden J, Altman DG, Gøtzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. International journal of epidemiology. 2007;36(4):847-57.

70. Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. Bmj. 2008;336(7644):601-5.
71. Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hróbjartsson A, et al. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. PLoS Med. 2006;3(10):e425.
72. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. Jama. 2001;285(15):2000-3.
73. Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. Neurology. 1994;44(1):16-20.
74. Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2000;49(3):359-70.
75. Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. J Eval Clin Pract. 2002;8(3):353-9.
76. Jørgensen KJ, Johansen HK, Gøtzsche PC. Flaws in design, analysis and interpretation of Pfizer's antifungal trials of voriconazole and uncritical subsequent quotations. Trials. 2006;7:3.
77. Savović J, Weeks L, Sterne JAC, Turner L, Altman DG, Moher D, et al. Evaluation of the Cochrane Collaboration’s tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. Systematic Reviews. 2014;3(1):37.
78. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898.
79. Babic A, Pijuk A, Brázdilová L, Georgieva Y, Raposo Pereira MA, Poklepovic Pericic T, et al. The judgement of biases included in the category "other bias" in Cochrane systematic reviews of interventions: a systematic survey. BMC Med Res Methodol. 2019;19(1):77.
80. Babic A, Tokalic R, Amílcar Silva Cunha J, Novak I, Suto J, Vidak M, et al. Assessments of attrition bias in Cochrane systematic reviews are highly inconsistent and thus hindering trial comparability. BMC Med Res Methodol. 2019;19(1):76.

81. Babic A, Vuka I, Saric F, Proloscic I, Slapnicar E, Cavar J, et al. Overall bias methods and their use in sensitivity analysis of Cochrane reviews were not consistent. J Clin Epidemiol. 2020;119:57-64.
82. Barcot O, Boric M, Dosenovic S, Poklepovic Pericic T, Cavar M, Puljak L. Risk of bias assessments for blinding of participants and personnel in Cochrane reviews were frequently inadequate. J Clin Epidemiol. 2019;113:104-13.
83. Barcot O, Boric M, Poklepovic Pericic T, Cavar M, Dosenovic S, Vuka I, et al. Risk of bias judgments for random sequence generation in Cochrane systematic reviews were frequently not in line with Cochrane Handbook. BMC Medical Research Methodology. 2019;19(1):170.
84. Propadalo I, Tranfic M, Vuka I, Barcot O, Pericic TP, Puljak L. In Cochrane reviews, risk of bias assessments for allocation concealment were frequently not in line with Cochrane's Handbook guidance. J Clin Epidemiol. 2019;106:10-7.
85. Puljak L, Ramic I, Arriola Naharro C, Brezova J, Lin YC, Surdila AA, et al. Cochrane risk of bias tool was used inadequately in the majority of non-Cochrane systematic reviews. J Clin Epidemiol. 2020;123:114-9.
86. Saric F, Barcot O, Puljak L. Risk of bias assessments for selective reporting were inadequate in the majority of Cochrane reviews. J Clin Epidemiol. 2019;112:53-8.
87. Barcot O, Dosenovic S, Boric M, Pericic TP, Cavar M, Jelicic Kadic A, et al. Assessing risk of bias judgments for blinding of outcome assessors in Cochrane reviews. Journal of comparative effectiveness research. 2020;9(8):585-93.
88. Barcot O, Boric M, Dosenovic S, Cavar M, Jelicic Kadic A, Poklepovic Pericic T, et al. Adequacy of risk of bias assessment in surgical vs non-surgical trials in Cochrane reviews: a methodological study. BMC Med Res Methodol. 2020;20(1):240.
89. Higgins JPT, Green S, (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration 2011. Available from http://handbook-5-1.cochrane.org/. Last accessed 22.11.2018.
90. Montori VM, Bhandari M, Devereaux PJ, Manns BJ, Ghali WA, Guyatt GH. In the dark: the reporting of blinding status in randomized controlled trials. J Clin Epidemiol. 2002;55(8):787-90.
91. Meinert CL. Masked monitoring in clinical trials--blind stupidity? The New England journal of medicine. 1998;338(19):1381-2.

92. Karanicolas PJ, Farrokhyar F, Bhandari M. Practical tips for surgical research: blinding: who, what, when, why, how? Canadian journal of surgery Journal canadien de chirurgie. 2010;53(5):345-8.
93. Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. J Clin Epidemiol. 2004;57(6):543-50.
94. Moustgaard H, Bello S, Miller FG, Hrobjartsson A. Subjective and objective outcomes in randomized clinical trials: definitions differed in methods publications and were often absent from trial reports. J Clin Epidemiol. 2014;67(12):1327-34.
95. Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration. Version 5.3. ed. Copenhagen: The Nordic Cochrane Centre: The Cochrane Collaboration; 2014.
96. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014;12(12):1495-9.
97. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt. 2014;34(5):502-8.
98. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. 1960;20(1):37-46.
99. Dechartres A, Trinquart L, Atal I, Moher D, Dickersin K, Boutron I, et al. Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. Bmj. 2017;357:j2490.
100. Kahan BC, Rehal S, Cro S. Risk of selection bias in randomised trials. Trials. 2015;16(1):405.
101. Savović J, Jones H, Altman D, Harris R, Jűni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. Health technology assessment (Winchester, England). 2012;16(35):1-82.
102. Hartling L, Hamm MP, Milne A, Vandermeer B, Santaguida PL, Ansari M, et al. Testing the Risk of Bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. J Clin Epidemiol. 2013;66(9):973-81.
103. da Costa BR, Hilfiker R, Egger M. PEDro's bias: summary quality scores should not be used in meta-analysis. J Clin Epidemiol. 2013;66(1):75-7.

104. da Costa BR, Beckett B, Diaz A, Resta NM, Johnston BC, Egger M, et al. Effect of standardized training on the reliability of the Cochrane risk of bias assessment tool: a prospective study. Systematic reviews. 2017;6(1):44.
105. Okuda PMM, Klaiman C, Bradshaw J, Reid M, Cogo-Moreira H. Assessing Risk of Bias in Randomized Controlled Trials for Autism Spectrum Disorder. Frontiers in psychiatry. 2017;8:265.
106. Rodrigues-Tartari R, Swardfager W, Salum GA, Rohde LA, Cogo-Moreira H. Assessing risk of bias in randomized controlled trials of methylphenidate for children and adolescents with attention deficit hyperactivity disorder (ADHD). International journal of methods in psychiatric research. 2018;27(1).
107. Cambridge Dictionary [Internet]. Cambridge University Press; 2019 [cited 15 August 2019]. Available from: https://dictionary.cambridge.org/dictionary/english/objective.
108. de Jong JD, Westert GP, Lagoe R, Groenewegen PP. Variation in hospital length of stay: do physicians adapt their length of stay decisions to what is usual in the hospital where they work? Health Serv Res. 2006;41(2):374-94.
109. Berger VW, Exner DV. Detecting selection bias in randomized clinical trials. Control Clin Trials. 1999;20(4):319-27.
110. Mickenautsch S, Fu B, Gudehithlu S, Berger VW. Accuracy of the Berger-Exner test for detecting third-order selection bias in randomised controlled trials: a simulation-based investigation. BMC Med Res Methodol. 2014;14:114.
111. Berger V. Selection Bias and Covariate Imbalances in Randomized Clinical Trials: John Wiley & Sons, Chichester; 2005.
112. Karanicolas PJ, Bickenbach K, Jayaraman S, Pusic AL, Coit DG, Guyatt GH, et al. Measurement and interpretation of patient-reported outcomes in surgery: an opportunity for improvement. J Gastrointest Surg. 2011;15(4):682-9.
113. Gurusamy KS, Gluud C, Nikolova D, Davidson BR. Assessment of risk of bias in randomized clinical trials in surgery. Br J Surg. 2009;96(4):342-9.

BIOGRAFIJA

OGNJEN BARČOT, DR. MED.

Curriculum vitae

Obrazovanje

1999–2005 dr. med. Medicinski fakultet, Sveučilište u Zagrebu

Edukacije

2018 Polaznik, Poslijediplomski doktorski studij TRIBE, MEF Split, HR
2014–2017 Uža specijalizacija abdominalne (digestivne) kirurgije, KBC Split, Split, HR
2008–2013 Specijalizacija opće kirurgije, KBC Split, Split, HR
2012–2014 Polaznik, Poslijediplomski doktorski studij Biologija Novotvorina, MEF Split
2011 Instruktor ALS – napredno održavanje života

Radno iskustvo

2018– Voditelj odjela abdominalne-koloproktološke kirurgije, KBC Split, Split
2017– Uži specijalist abdominalne (digestivne) kirurgije, KBC Split, Split
2014–2016 Tajnik Hrvatskog društva za endoskopsku kirurgiju (HDEK)
2013–2017 Specijalist opće kirurgije, Klinika za kirurgiju, KBC Split, Split
2008–2013 Specijalizacija opće kirurgije, Klinika za kirurgiju, KBC Split, Split
2006–2008 Liječnik u timu hitne medicinske pomoći, UHMP SDŽ, Split
2006 Liječnik u praksi obiteljske medicine, DZ SDŽ, Split

Osobne vještine

Strani jezik – engleski – razina C1/C2
Digitalne vještine – ECDL Start certifikat (M2, M3, M4, M6, M7)

Priznanja i nagrade

2004 Rektorova nagrada Sveučilišta u Zagrebu

Članstva

2013– HDEK - hrvatsko društvo za endoskopsku kirurgiju
2013– EAES - europsko društvo za endoskopsku kirurgiju

ZNANSTVENE PUBLIKACIJE:

U razdoblju od 01/01/2007. do 05/02/2021 objavio 11 radova u recenziranim znanstvenim časopisima.
Broj citata na datum 05/02/2021:
Google Scholar: broj citata 73, h-indeks: 5, i10-indeks: 3

1: **Barcot O**, Boric M, Cavar M, Poropat G, Puljak L. Predictive value of abdominal drain amylase for postoperative pancreatic fistula following gastric resection: A systematic review and meta-analysis. Eur J Surg Oncol. 2020 Dec 13:S0748-7983(20)31224-5. doi: 10.1016/j.ejso.2020.12.005. Epub ahead of print. PMID: 33334630.

2: **Barcot O**, Boric M, Dosenovic S, Cavar M, Jelicic Kadic A, Poklepovic Pericic, T, Vukicevic I, Vuka I, Puljak L. Adequacy of risk of bias assessment in surgical vs non-surgical trials in Cochrane reviews: a methodological study. BMC Med Res Methodol. 2020 Sep 29;20(1):240. doi: 10.1186/s12874-020-01123-7. PMID: 32993499; PMCID: PMC7526117.

3: Augustin G, Boric M, **Barcot O**, Puljak L. Discordant outcomes of laparoscopic versus open appendectomy for suspected appendicitis during pregnancy in published meta-analyses: an overview of systematic reviews. Surg Endosc. 2020 Oct;34(10):4245-4256. doi: 10.1007/s00464-020-07674-6. Epub 2020 Jun 16. PMID: 32556754.

4: **Barcot O**, Dosenovic S, Boric M, Pericic TP, Cavar M, Jelicic Kadic A, Puljak L. Assessing risk of bias judgments for blinding of outcome assessors in Cochrane reviews. J Comp Eff Res. 2020 Jun;9(8):585-593. doi: 10.2217/cer-2019-0181. Epub 2020 May 27. PMID: 32459105.

5: Puljak L, Ramic I, Arriola Naharro C, Brezova J, Lin YC, Surdila AA, Tomajkova E, Farias Medeiros I, Nikolovska M, Poklepovic Pericic T, **Barcot O**, Suarez Salvado M. Cochrane risk of bias tool was used inadequately in the majority of non-Cochrane systematic reviews. J Clin Epidemiol. 2020 Jul;123:114-119. doi: 10.1016/j.jclinepi.2020.03.019. Epub 2020 Apr 1. PMID: 32247026.

6: Könsgen N, **Barcot O**, Heß S, Puljak L, Goossen K, Rombey T, Pieper D. Inter-review agreement of risk-of-bias judgments varied in Cochrane reviews. J Clin Epidemiol. 2020 Apr;120:25-32. doi: 10.1016/j.jclinepi.2019.12.016. Epub 2019 Dec 19. PMID: 31866473.

7: **Barcot O**, Boric M, Poklepovic Pericic T, Cavar M, Dosenovic S, Vuka I, Puljak L. Risk of bias judgments for random sequence generation in Cochrane systematic reviews were frequently not in line with Cochrane Handbook. BMC Med Res Methodol. 2019 Aug 5;19(1):170. doi: 10.1186/s12874-019-0804-y. PMID: 31382898; PMCID: PMC6683577.

8: **Barcot O**, Boric M, Dosenovic S, Poklepovic Pericic T, Cavar M, Puljak L. Risk of bias assessments for blinding of participants and personnel in Cochrane reviews were frequently inadequate. J Clin Epidemiol. 2019 Sep;113:104-113. doi: 10.1016/j.jclinepi.2019.05.012. Epub 2019 May 24. PMID: 31132470.

9: Saric F, **Barcot O**, Puljak L. Risk of bias assessments for selective reporting were inadequate in the majority of Cochrane reviews. J Clin Epidemiol. 2019 Aug;112:53-58. doi: 10.1016/j.jclinepi.2019.04.007. Epub 2019 Apr 19. PMID: 31009658.

10: Propadalo I, Tranfic M, Vuka I, **Barcot O**, Pericic TP, Puljak L. In Cochrane reviews, risk of bias assessments for allocation concealment were frequently not in line with Cochrane's Handbook guidance. J Clin Epidemiol. 2019 Feb;106:10-17. doi: 10.1016/j.jclinepi.2018.10.002. Epub 2018 Oct 9. PMID: 30312657.

11: **Barcot O**, Balarin M, Gamulin O, Jezek D, Romac P, Brnjas-Kraljević J. Investigation of spermatozoa and seminal plasma by fourier transform infrared spectroscopy. Appl Spectrosc. 2007 Mar;61(3):309-13. doi: 10.1366/000370207780220804. PMID: 17389071.