

Maske ne obvarujejo pred virusi. Ali pač?

Avtomatsko ugotavljanje kakovosti izsledkov v medicini

Simon Šuster
Seminar JOTA
5. junij 2023





Objave

[1] Šuster S, Baldwin T, Lau, JH, Jimeno Yepes A, Martinez Iraola D, Otmakhova Y, Verspoor K. (2023) Automating Quality Assessment of Medical Evidence in Systematic Reviews. *Journal of Medical Internet Research* 2023, e35568. DOI: [10.2196/35568](https://doi.org/10.2196/35568)

[2] Šuster S, Baldwin T, Verspoor K*. (2023) Analysis of predictive performance and reliability of classifiers for quality assessment of medical evidence revealed important variation by medical area. *Journal of Clinical Epidemiology*. DOI: [10.1016/j.jclinepi.2023.04.006](https://doi.org/10.1016/j.jclinepi.2023.04.006)

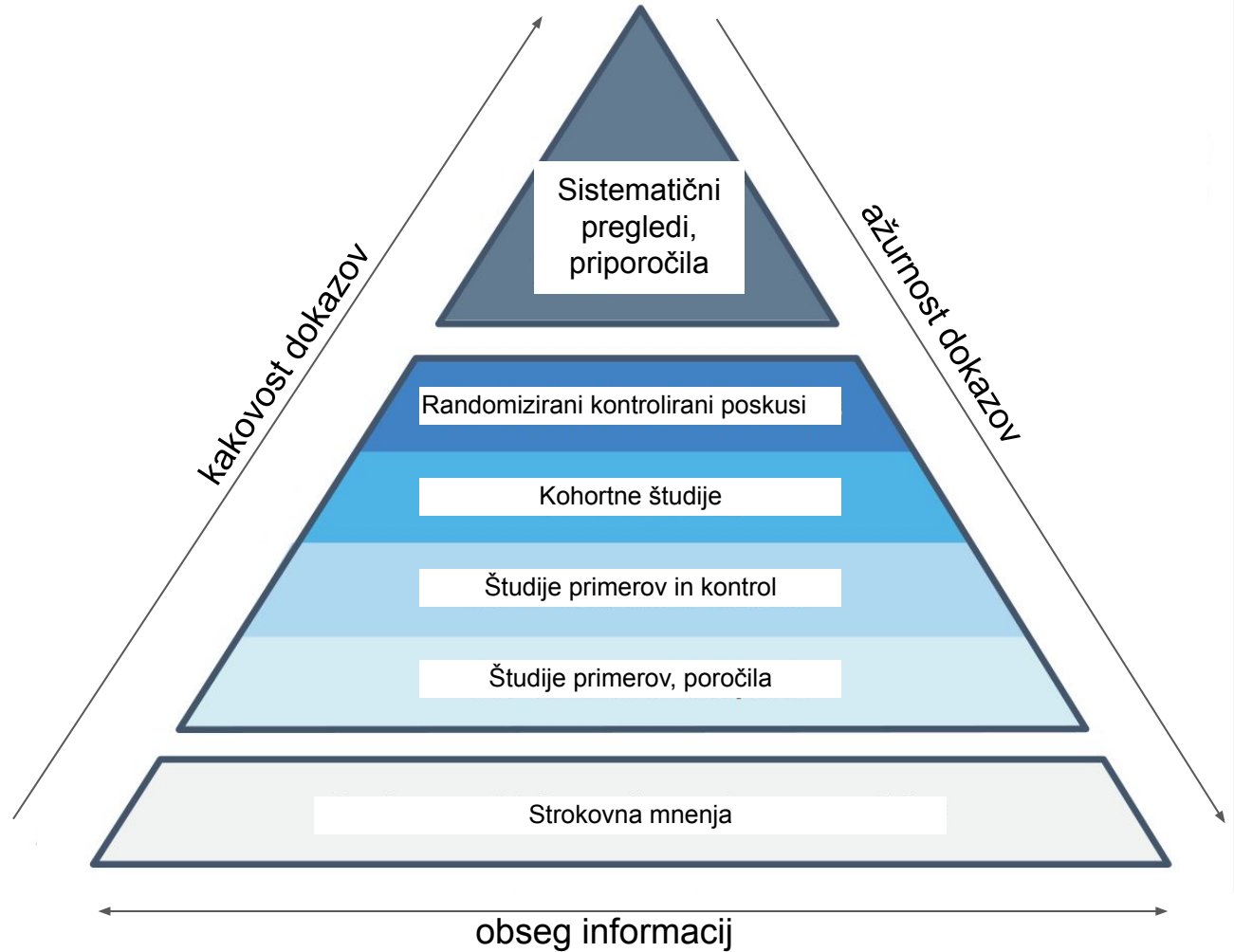
[3] Šuster S, Baldwin T, Verspoor K. (2023) Does Promoting Fairness in Classification of Quality of Medical Evidence Work? *BioNLP, ACL*.

Na dokazih temelječa medicina (Evidence-based medicine)

Klinično odločanje naj temelji na znanstvenih izsledkih, vendar:

1. znanstveni izsledki sami še ne zadoščajo za klinično odločanje (vrednotenje)
2. postavljeni morajo biti v ustrezno hierarhijo

Ravni dokazov



Physical interventions to interrupt or reduce the spread of respiratory viruses

Tom Jefferson, Liz Dooley, Eliana Ferroni, Lubna A Al-Ansary, Mieke L van Driel, Ghada A Bawazeer, Mark A Jones, Tammy C Hoffmann, Justin Clark, Elaine M Beller, Paul P Glasziou, ✉ John M Conly Authors' declarations of interest

Version published: 30 January 2023 Version history

<https://doi.org/10.1002/14651858.CD006207.pub6> 

Glavni izsledek:

→”rezultati randomiziranih kontroliranih poskusov (RKP) niso pokazali jasnega zmanjšanja respiratornih virusnih okužb zaradi uporabe medicinskih/kirurških mask.”

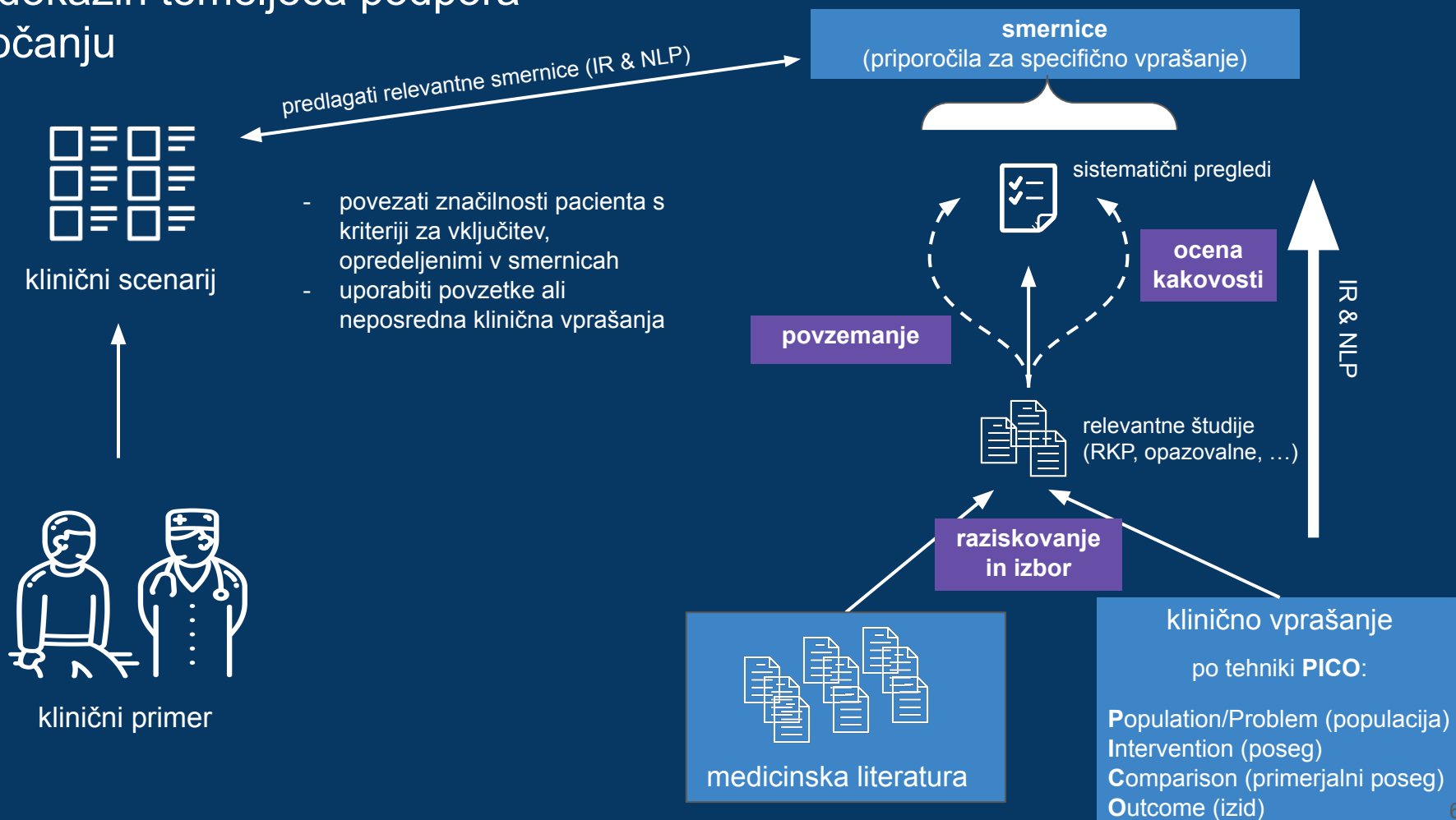
Obrazložitev:

→”obstaja negotovost glede učinkov, ki jih ima nošenje obraznih mask. Nizka do srednja dokazna gotovost pomeni, da je naše zaupanje v ocenjen učinek omejeno in da se ta v resnici lahko razlikuje”

→”Zaradi visokega tveganja za pristranskost v poskusih, razlik v merjenju izidov in relativne nizke sodelovalnosti ni mogoče priti do trdnih zaključkov”

→”potrebni je več kakovostnih RKP, s katerimi bi lahko ovrednotili najbolj učinkovite strategije za uspešno izvajanje fizičnih posegov v praksi [...] več prizadevanj, strožje načrtovanje in povečano financiranje za visokokakovostne RKP žal niso bili zagotovljeni v trenutni pandemiji covid-19 [...] Več poudarka bi moralo biti na opisovanju in oceni škod, nastalih zaradi posegov, vključenih v ta pregled, ter na njihovem razmerju s sodelovalnostjo.”

Na dokazih temelječa podpora odločanju



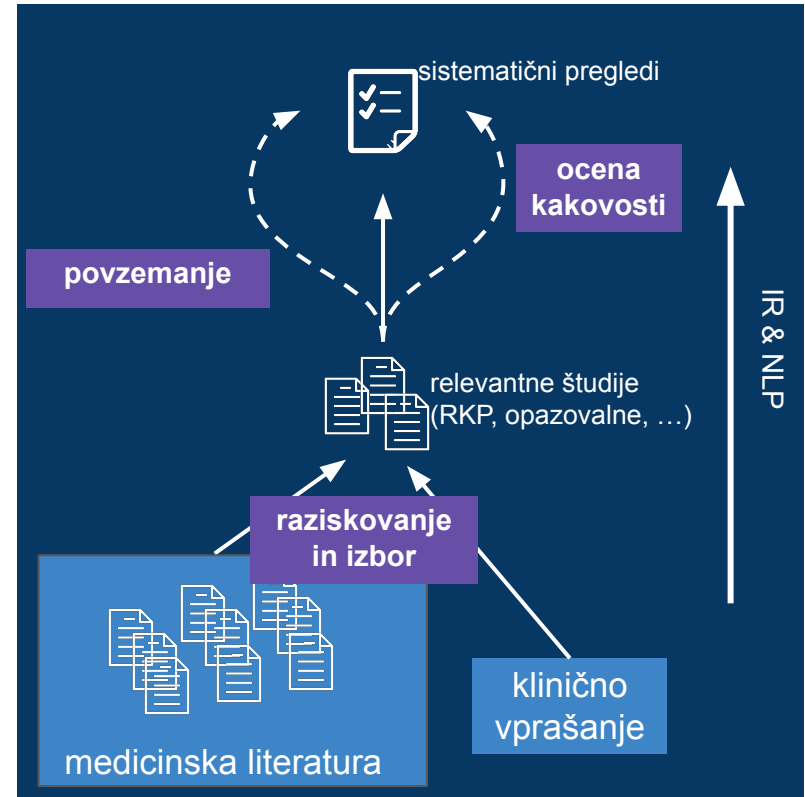
Zakaj avtomatizirati

Sistematični pregledi hitro zastarajo:

- izredno zamudna izdelava (1–2 leti)
- naraščajoče število objav primarnih študij

Kako pospešiti proces izdelave s strojnim učenjem:

- doslej poudarek na iskanju in presejanju
- ocenjevanje kakovosti omejeno na zaznavanje tveganja za pristranskost v posamičnih študijah



Mesečni prirast člankov o covidu-19

LitCovid Data

352,326

Publications

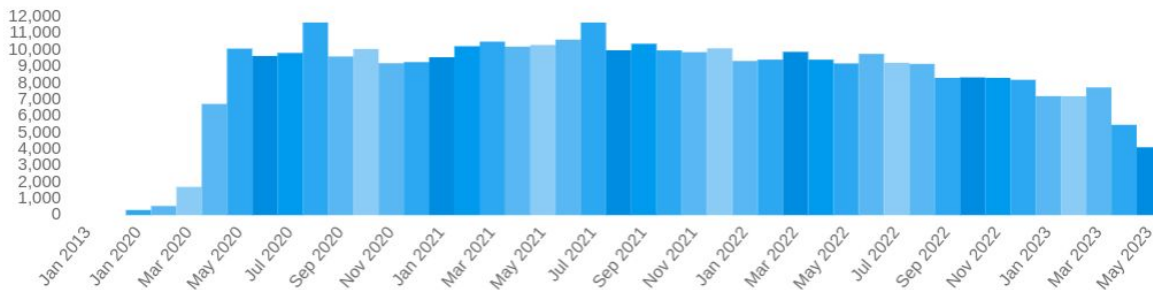
8

Topics

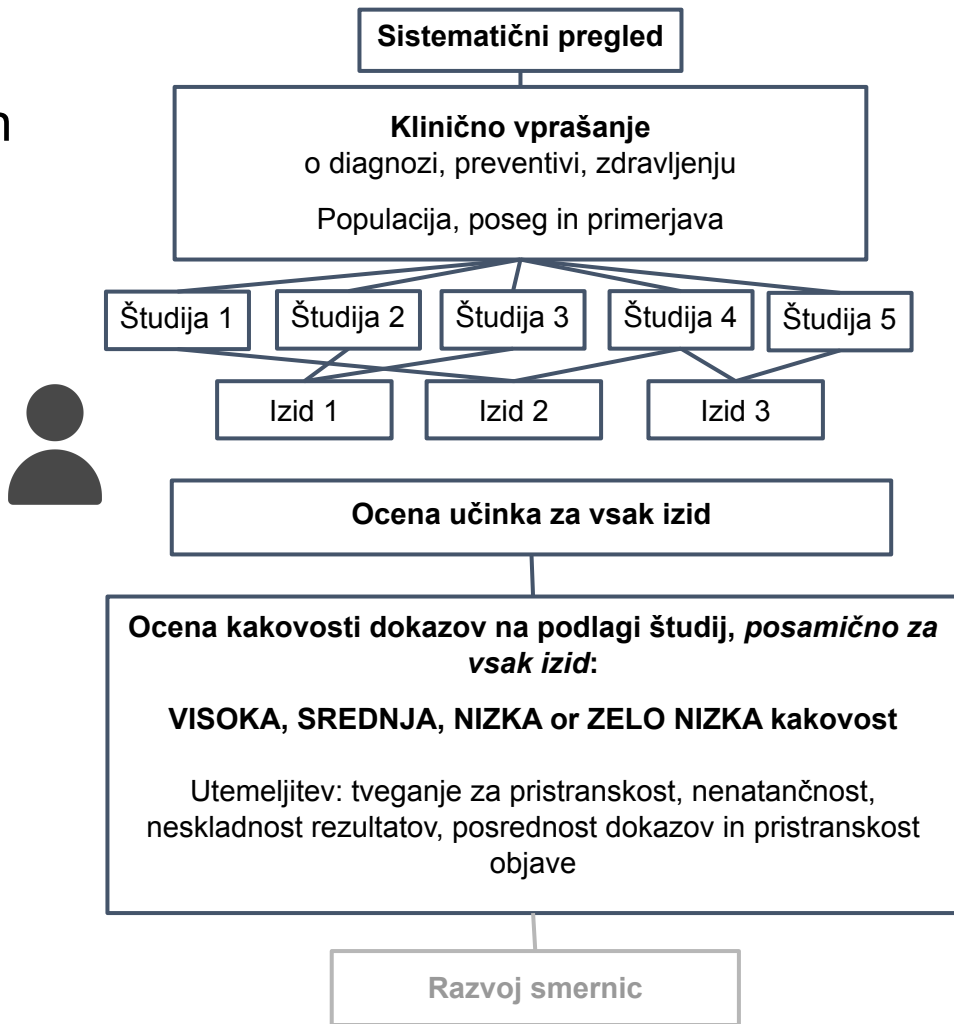
8,000

Journals

Monthly Publications

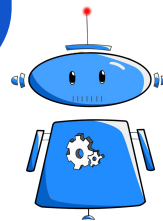


Priprava sistematičnih pregledov in ocenjevanje kakovosti



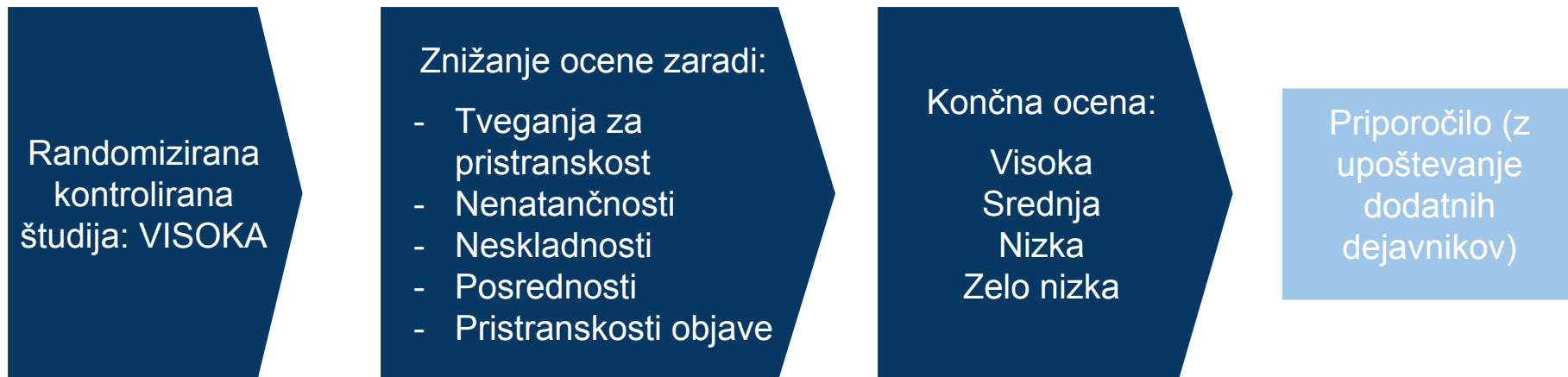
Naš cilj (EvidenceGRADER):

Na podlagi skupka dokazov napovedati njihovo kakovost



Podatkovna množica + naloge + modeli s heterogenimi vhodnimi podatki (strukturiranimi in nestrukturiranimi)

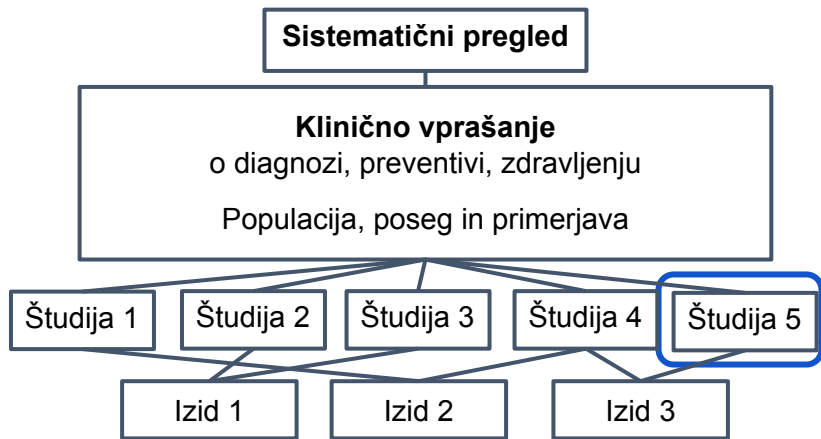
Kakovost dokazov po protokolu GRADE (Grading of Recommendations Assessment, Development and Evaluation)



Preglednica 5: Pristop GRADE – kakovost dokazov in glavni kriteriji za določanje jakosti priporočil.

Kakovost dokazov		
++++	visoka	nadaljnje raziskave najverjetneje ne bodo spremenile našega zaupanja v ocenjeni učinek = prepričani smo, da je dejanski učinek v bližini ocenjenega
+++	srednja	nadaljnje raziskave bodo verjetno pomembno vplivale na naše zaupanje v ocenjeni učinek in lahko oceno spremenijo = dejanski učinek je najbrž blizu ocenjenemu, možno pa je tudi, da se od njega bistveno razlikuje
++	nizka	nadaljnje raziskave bodo zelo verjetno pomembno vplivale na naše zaupanje v ocenjeni učinek in bodo oceno verjetno spremenile = dejanski učinek se lahko bistveno razlikuje od ocenjenega
+	zelo nizka	kakršnakoli ocena učinka je zelo negotova
Jakost priporočil bodisi močna bodisi šibka – glede na		
		različen pomen različnih izidov (večjega naj imajo pomembnejši)
		razmerje med želenimi in neželenimi učinki zdravljenja
		enotnost (gotovost) oziroma raznolikost (negotovost) vrednot in izbor zdravljenja
		stroške (potrebna sredstva)

Zaznavanje tveganja za pristranskost v sklopu sistematičnih pregledov



Sorodni sistemi:



Na podlagi izvlečka oz. celotnega članka oceniti tveganje za pristranskost

Ocena tveganja za pristranskost za *posamično študijo*

- ustvarjanje naključnega niza (*rand. seq. generation*)
- zakrivanje informacij o uvrščanju (*allocation concealment*),
- slepi načrt raziskav (*blinding of participants & personnel, blinding of outcome assessment*)

Končna ocena:
NIZKO, VISOKO/NEJASNO
+ utemeljitev s fragmentom članka

Tabele dokazov: pridobivanje oznak/pravih vrednosti in strukturiranih značilik

Levosimendan compared with placebo for cardiogenic shock or low cardiac output syndrome

Patient or population: adults with cardiogenic shock or low cardiac output syndrome

Settings: hospital

Intervention: levosimendan

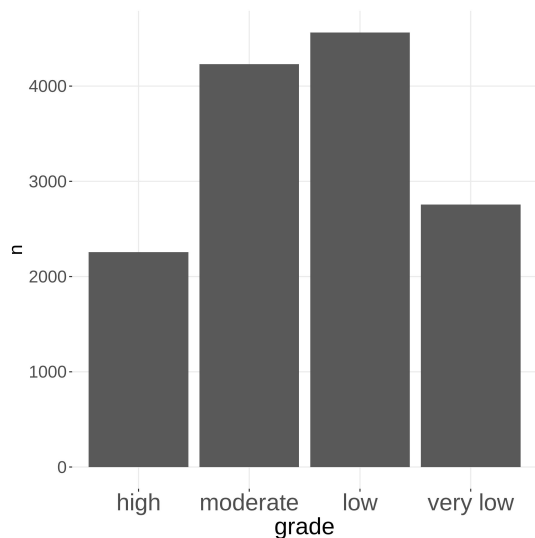
Comparison: placebo

Outcomes	Anticipated absolute effects (95% CI)		Relative effect (95% CI)	No of participants (studies)	Quality	Comments
	Risk with placebo	Risk with levosimendan				
All-cause short-term mortality: range 4 to 6 months	Moderate¹		RR 0.48 (0.12 to 1.94)	55 (2)	⊕⊕⊕⊖ very low ^{3,4}	Studies included participants with LCOS or CS due to HF or AMI
	187 per 1000	90 per 1000 (22 to 363)				
	High²					
	500 per 1000	240 per 1000 (60 to 970)				

³Downgraded one step due to study limitation because of lack of blinding of participants and physicians, and missing information on randomisation in the larger study.

⁴Downgraded two steps for imprecision due to few events and the confidence interval crosses the line of no difference and includes possible benefit from both approaches.

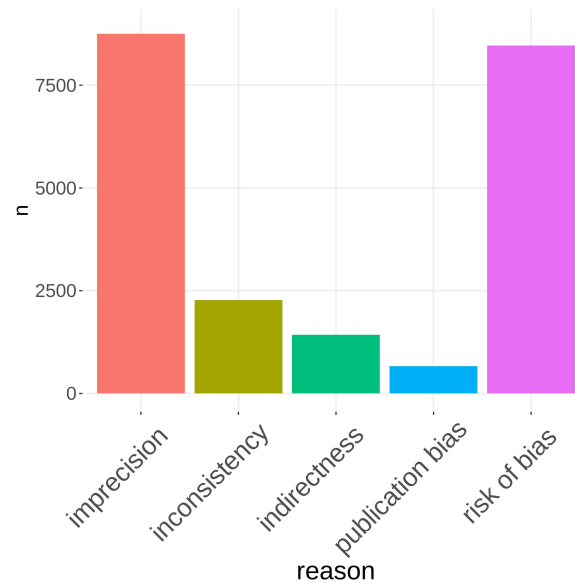
Porazdelitev podatkov



Kakovost dokazov je visoka le v 1/6 primerov


Polovica primerov je (zelo) nizke kakovosti

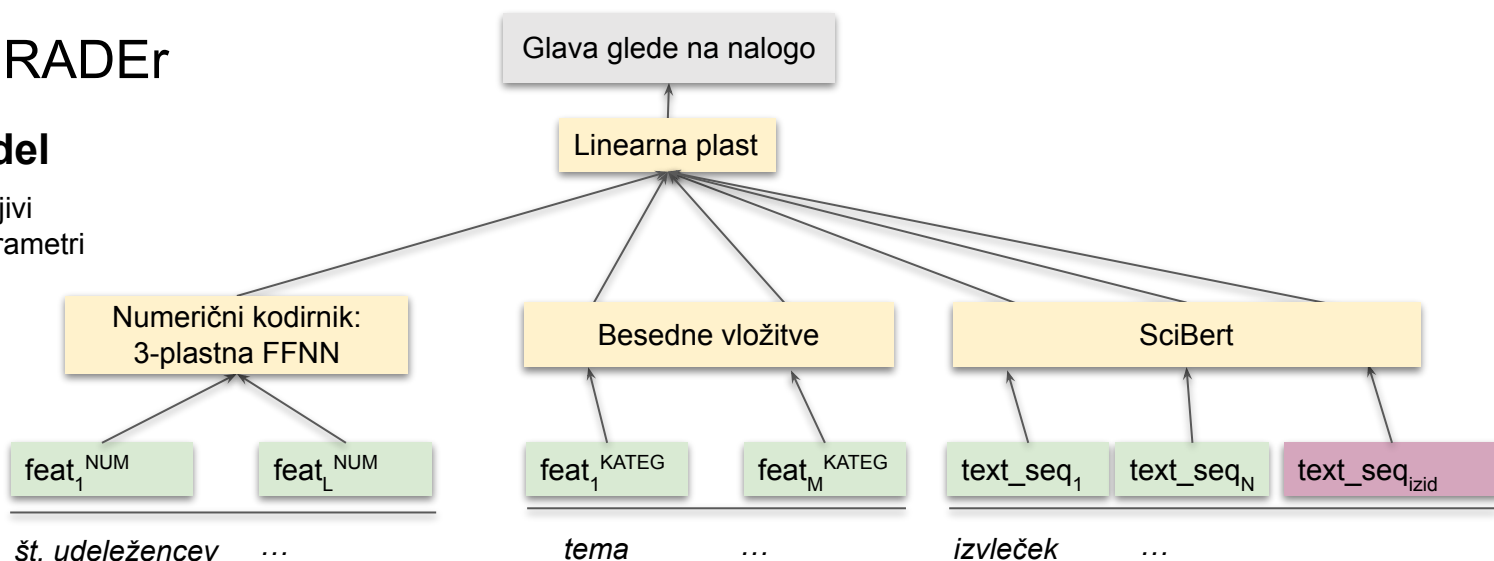
Ocena kakovosti dokazov se zniža zaradi:



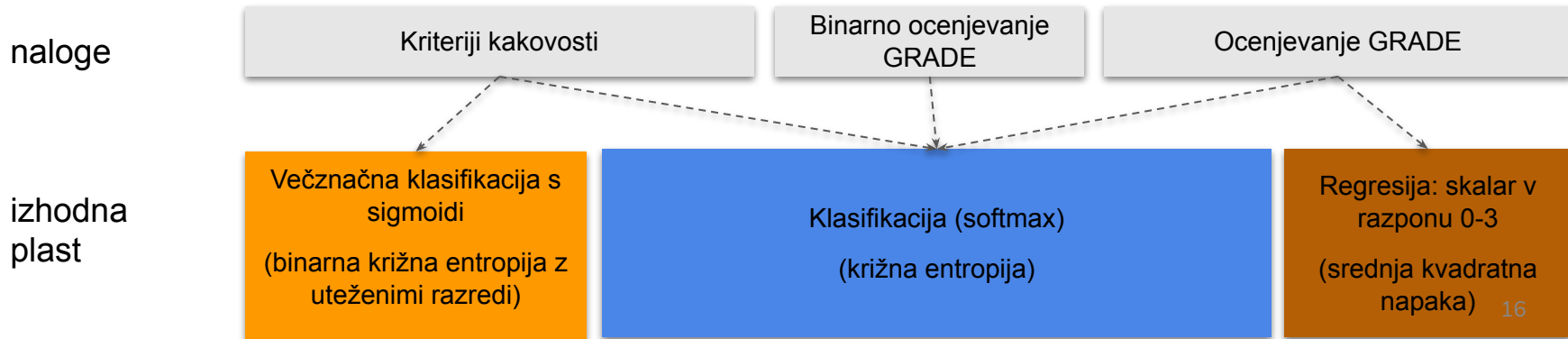
EvidenceGRADER

Osnovni model

 : učljivi parametri



Specifične naloge



Rezultati za EvidenceGRADER

(povprečne vrednosti iz 10-kratnega prečnega preverjanja)

Ocenjevanje GRADE

možnost
arbitrarnih
razlik

sorodni
sistem

	F1	MAE
EvidenceGRADER	0.49 (P: 0.54, R: 0.49)	0.62
naključna metoda	0.25	1.2
večinska metoda	0.13	0.8

RobotReviewer (naša reevalvacija)

Kriteriji kakovosti

	F1
nenatančnost (<i>imprecision</i>)	0.63
tveganje za pristranskost (<i>risk of bias</i>)	0.70
neskladnost (<i>inconsistency</i>)	0.21
posrednost (<i>indirectness</i>)	0.26
pristranskost objave (<i>publication bias</i>)	0.11

	F1
random seq. generation	72.2
allocation concealment	66.5
blinding of participants	74.0
blinding of outcome assessment	62.6

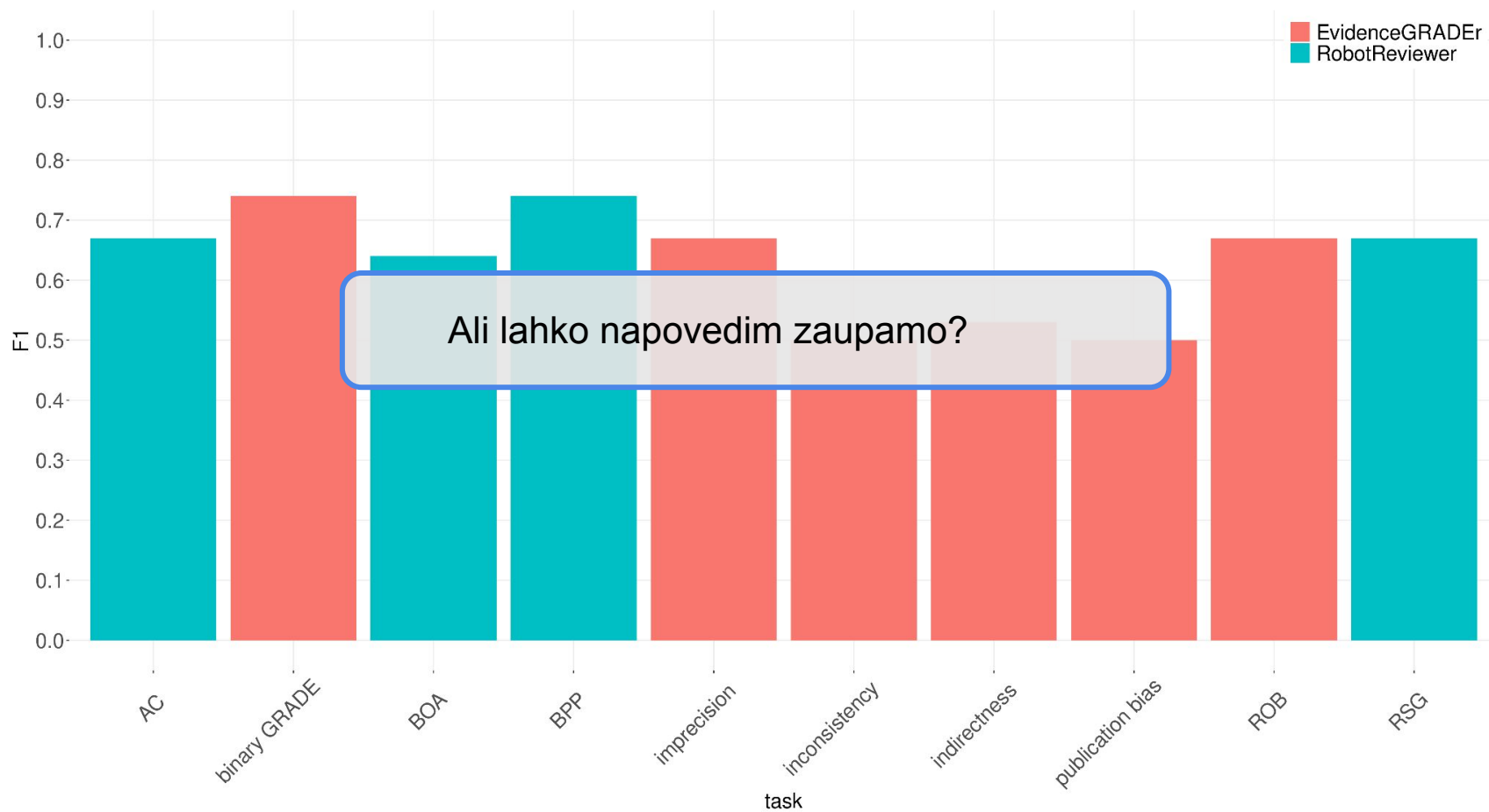
Binarno ocenjevanje GRADE

poenostavljena
naloga

	F1
EvidenceGRADER	0.74
naključna metoda	0.43
večinska metoda	0.46

določene
kategorije
so redke

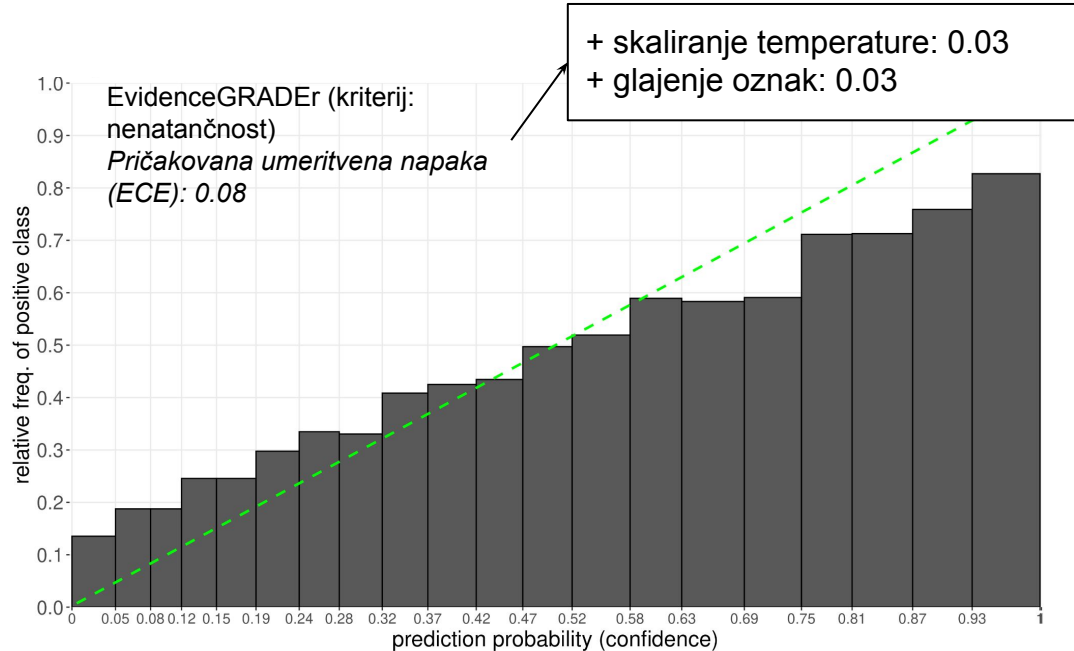
Uspešnost napovedovanja



Umerjanje klasifikatorjev

Pretirana gotovost nevronske mreže

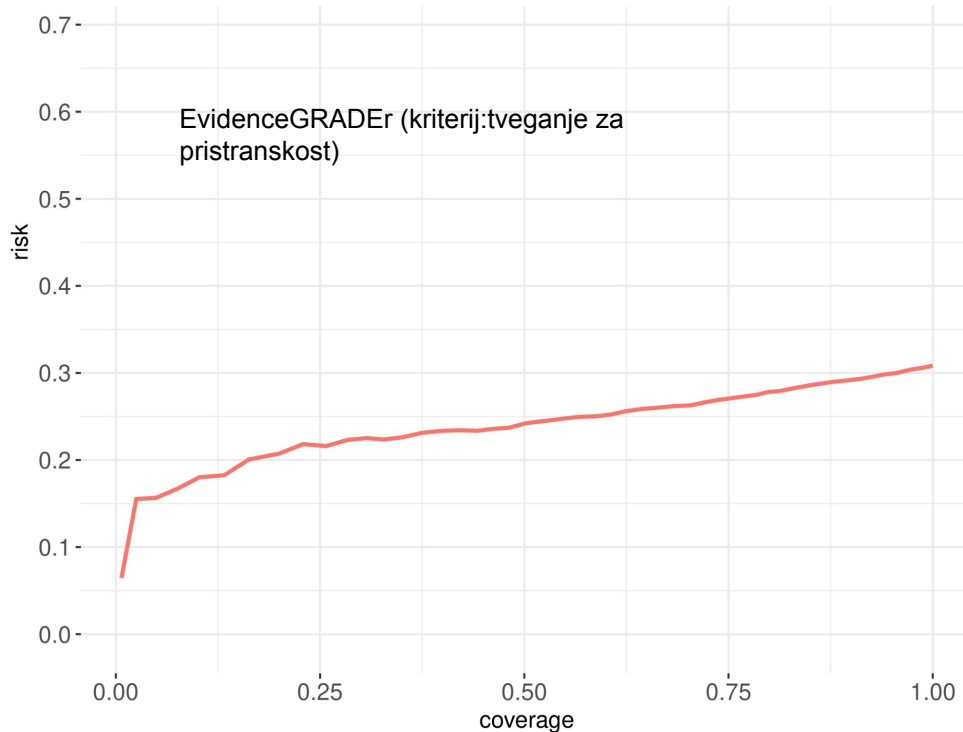
Model je zanesljiv, kadar ve, česar ne ve



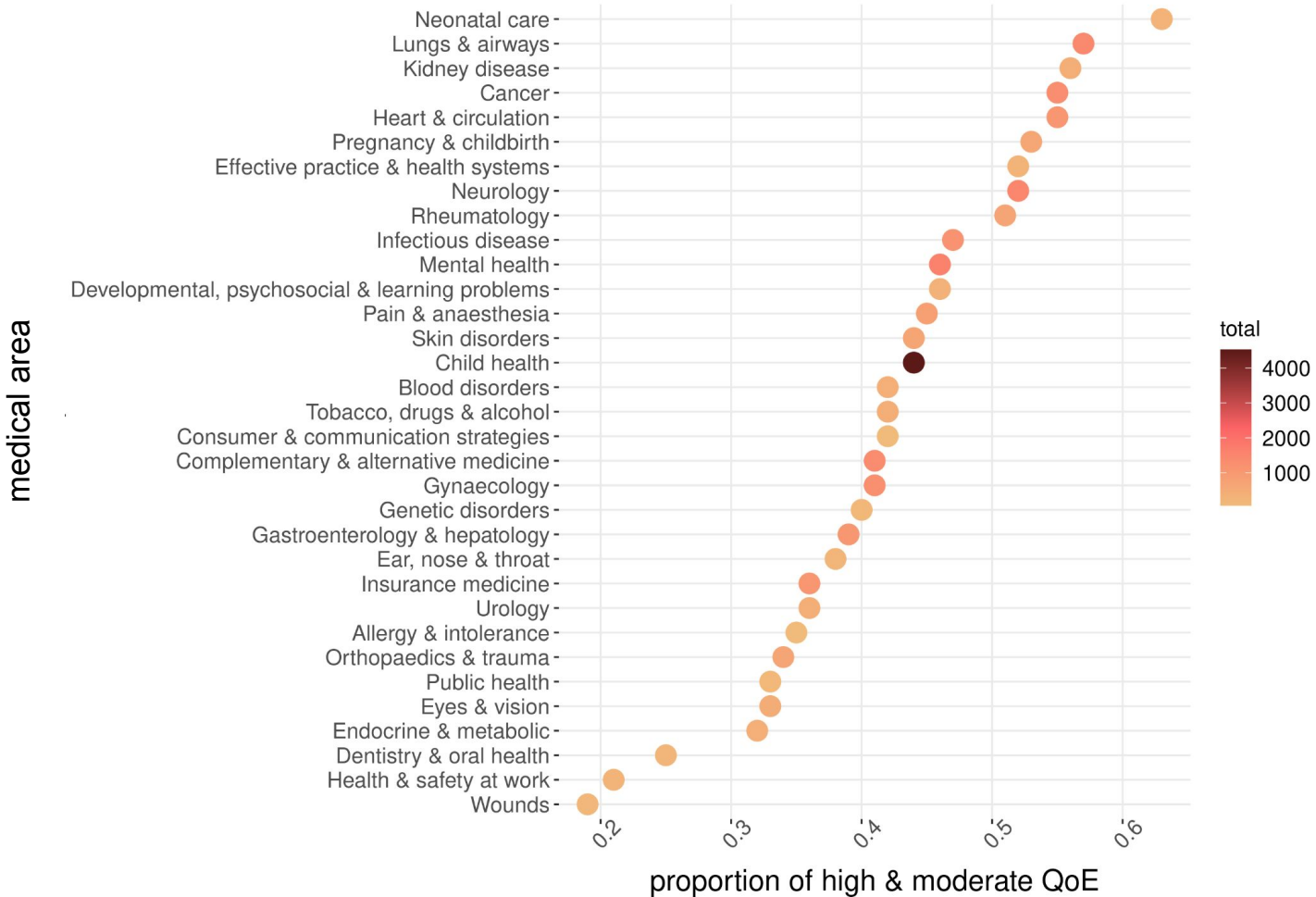
Izbirno napovedovanje (1/2)

Nekatere napovedi obdržimo, nekatere zavržemo

cilj: zmanjšati verjetje za napako ob hkratnem zmanjšanju pokritosti

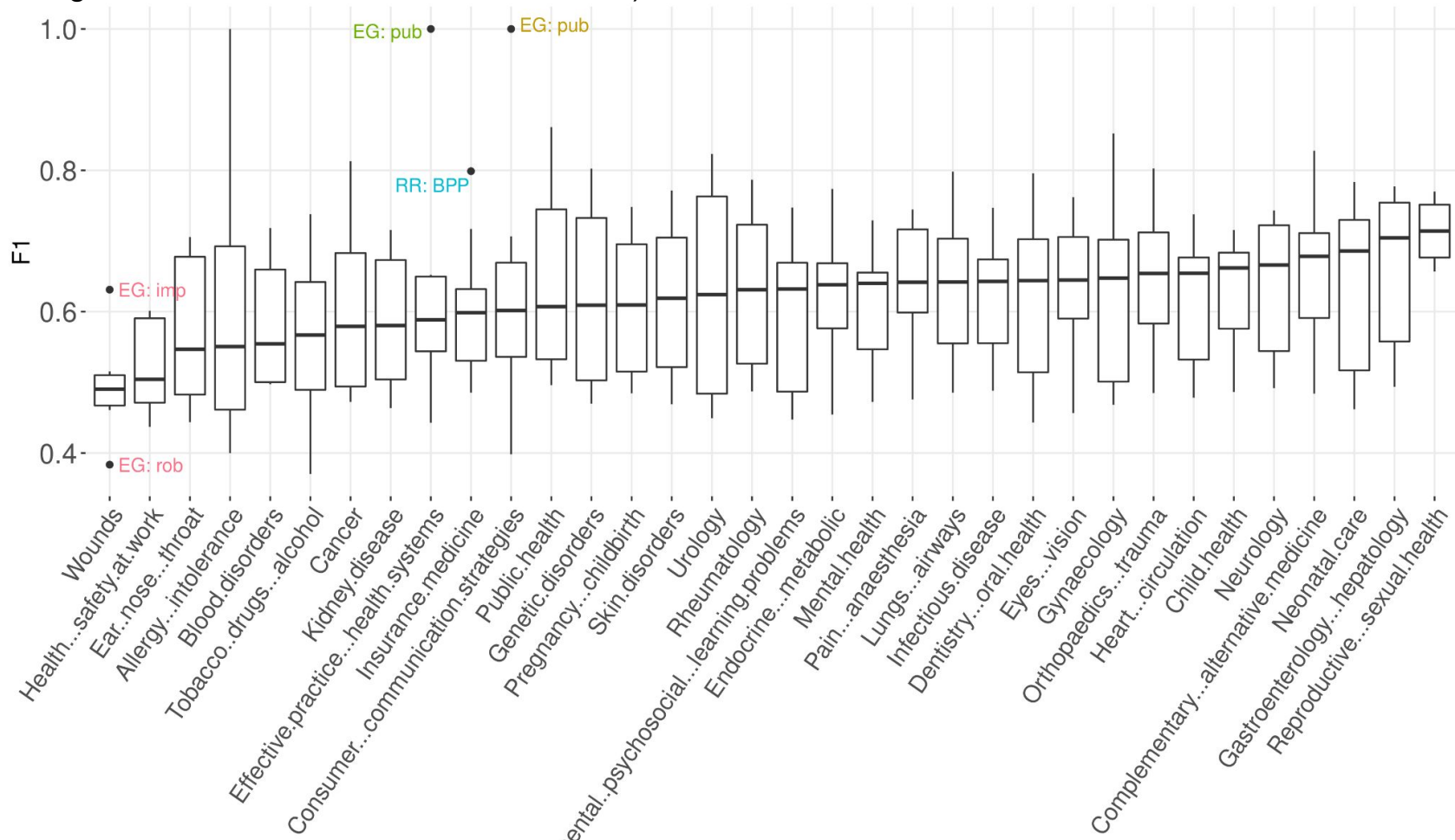


Nesorazmerje srednje-/visokokakovostnih dokazov v sist. pregledih Cochrane



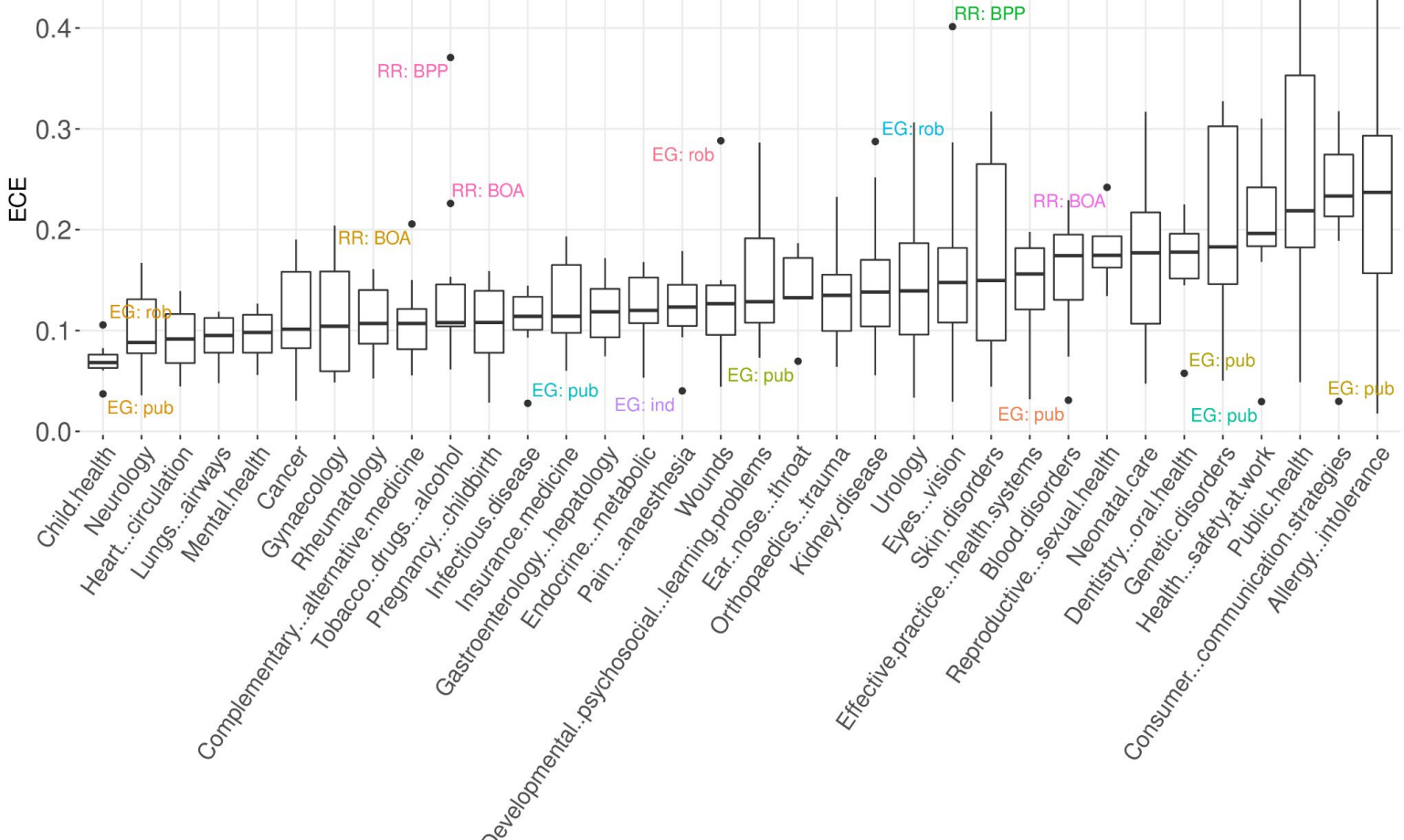
Neenakost v meri uspešnosti F1 po področjih

(vse naloge EvidenceGRADER & RobotReviewer)



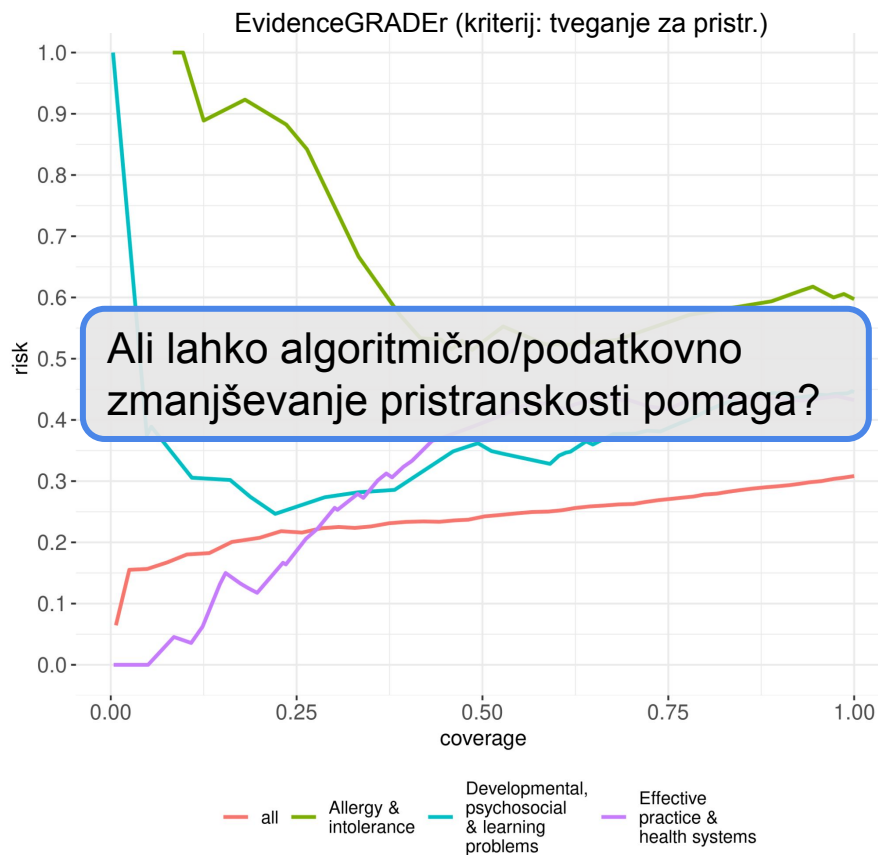
Neenakost v meri zanesljivosti ECE po področjih

(vse naloge EvidenceGRADER & RobotReviewer)

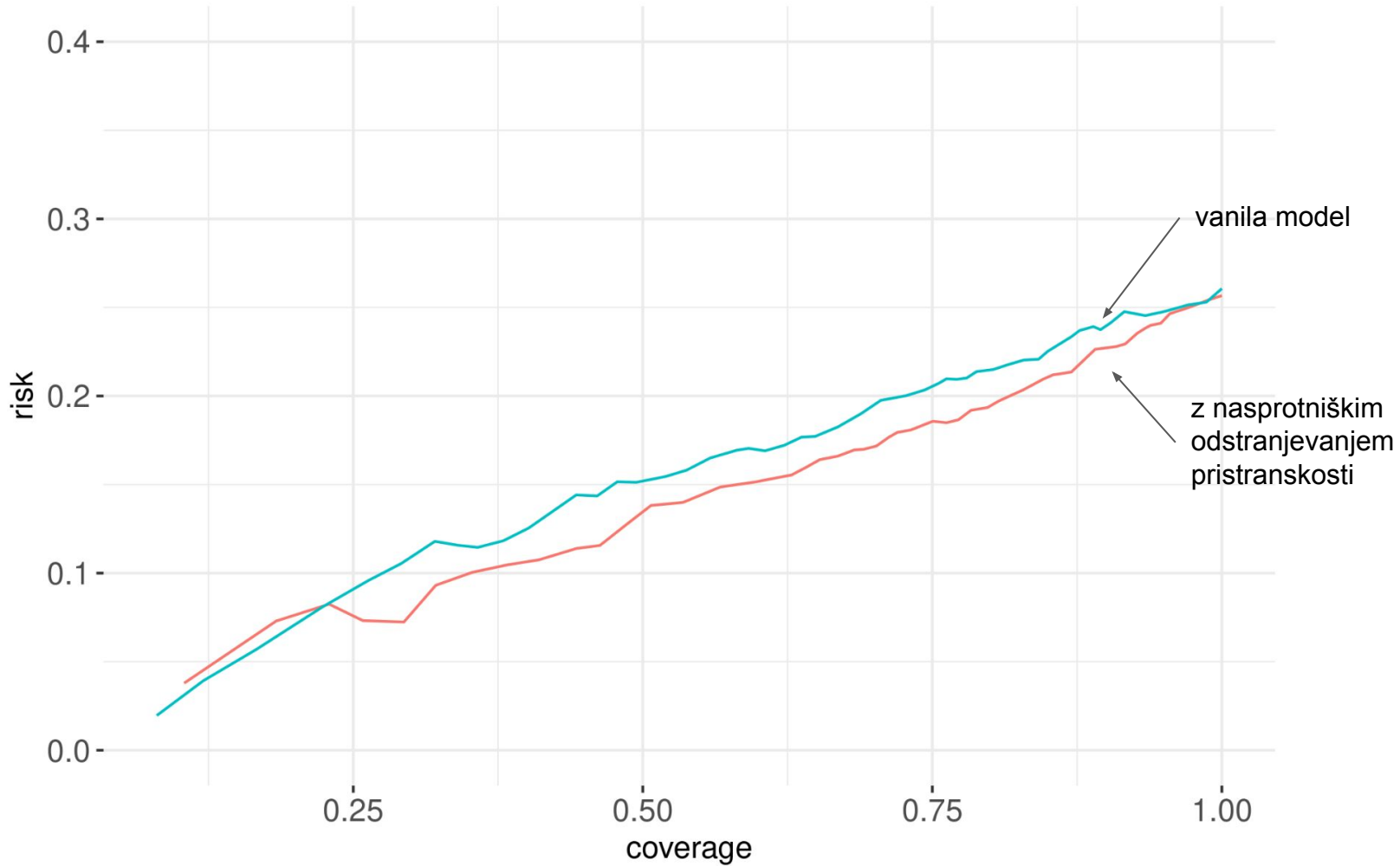


Izbirno napovedovanje (2/2)

povprečje proti trem področjem z najnižjo F1 pri polni pokritosti



binaryGRADE: Cancer



Poštenost (nepristranskost)

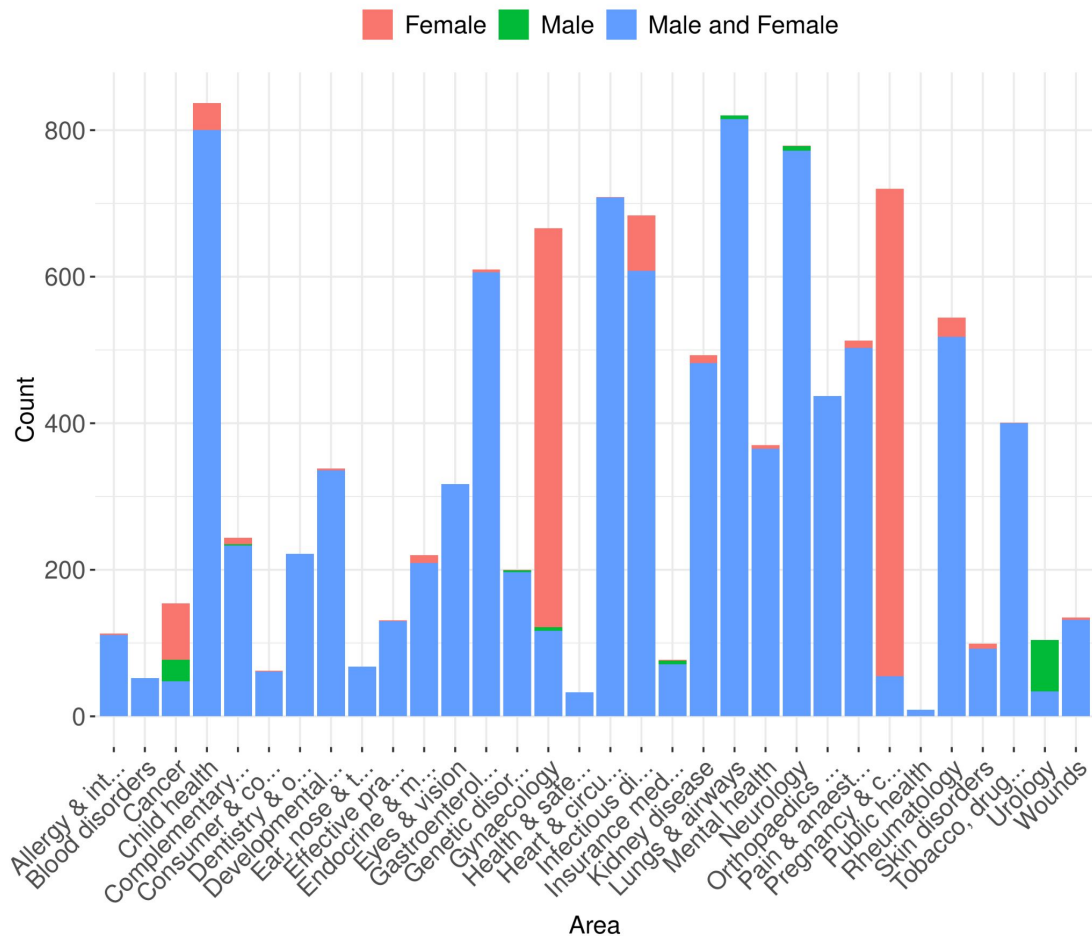
Poštenost glede na dva atributa:

- področje
- biološki spol

Vpliv na uspešnost:

- napovedovanja
- izbirne klasifikacije
- umerjenosti

Spodbujanje nepristranskosti po načelu enakih možnosti → kompromisi med merami uspešnosti



Vloga v praksi

Polavtomatsko, kot podpora pri ročnem delu:

- dodaten “označevalec” (preverjanje skladnosti)
- glede na gotovost in uspešnost klasifikatorja bolj ali manj avtonomno delovanje

Omejen nabor vhodnih značilk v določenih fazah gradnje sist. pregleda

Kombiniranje z obstoječim sistemom RobotReviewer

Zaključek

- napovedovanje kakovosti izsledkov
- priprava podatkovne množice iz obstoječih virov
- rezultati odvisni od naloge (4- ali 2-točkovno, posamezni kriteriji)
- pomen izbirne klasifikacije in umerjenosti za praktično uporabo
- neuravnoteženost dokazov in spodbujanje nepristranskosti