

UNIVERZITET U BEOGRADU

MEDICINSKI FAKULTET

Zoran M. Bukumirić

**SISTEMATSKI PREGLED SA META-ANALIZOM
POUZDANOSTI I VALJANOSTI PSIHIJATRIJSKIH
SKALA U LONGITUDINALNIM STUDIJAMA**

doktorska disertacija

Beograd, 2016

**UNIVERSITY OF BELGRADE
SCHOOL OF MEDICINE**

Zoran M. Bukumirić

**SYSTEMATIC REVIEW AND META-ANALYSIS OF
RELIABILITY AND VALIDITY OF PSYCHIATRIC
SCALES IN LONGITUDINAL STUDIES**

Doctoral Dissertation

Belgrade, 2016

Mentor:

Prof. dr Goran Trajković, Medicinski fakultet, Univerzitet u Beogradu

Članovi komisije:

Doc. dr Dejana Stanisavljević, Medicinski fakultet, Univerzitet u Beogradu

Doc. dr Nataša Milić, Medicinski fakultet, Univerzitet u Beogradu

Prof. dr Slavica Đukić Dejanović, Medicinski fakultet, Univerzitet u Kragujevcu

Datum odbrane: _____

*Izrada doktorske disertacije je proces koji često traje godinama
i zahteva direktno ili indirektno uključivanje i drugih.*

*Pre svega želim da izrazim zahvalnost svom mentoru prof. dr Goranu Trajkoviću
čije su znanje i iskustvo nepresušni izvor inspiracije i večita opomena
da uvek može i više i bolje. Hvala mu na zalaganju, posvećenosti i podršci
koja je obeležila naš zajednički rad.*

*Veliku zahvalnost dugujem članovima komisija doc. dr Dejani Stanisavljević,
doc. dr Nataši Milić, prof. dr Slavici Đukić Dejanović,
prof. dr Jeleni Marinković Erić i doc. dr Aniti Grgurević
na sugestijama u toku izrade ove disertacije.*

*Iskreno se zahvaljujem porodici, rodbini, prijateljima i kolegama
koji su me bodrili i imali razumevanja za sva moja odsustvovanja.*

*Neizmernu zahvalnost dugujem supruzi Dragici i ćerki Nađi, bez čije velike ljubavi
i razumevanja izrada ove doktorske disertacije ne bi bila moguća.*

*Ovu disertaciju posvećujem mojim dragim roditeljima, Mileti i Verici,
koji su nesebično uložili sebe u moje vaspitanje i obrazovanje.*

**SISTEMATSKI PREGLED SA META-ANALIZOM
POUZDANOSTI I VALJANOSTI PSIHIJATRIJSKIH SKALA
U LONGITUDINALNIM STUDIJAMA**

Bukumirić M. Zoran

Rezime:

Uvod: Psihijatrijske skale omogućavaju da se registruje prisustvo ili odsustvo simptoma psihijatrijskih oboljenja, da se njihov intenzitet izrazi brojevima i upotrebi u kliničke ili istraživačke svrhe. Samim tim, važan aspekt skala u psihijatriji jesu njihove metrijske osobine, odnosno njihova merna adekvatnost što obuhvata pouzdanost i valjanost. Pouzdanost označava do koje mere se metoda merenja (test) konzistentno ili postojano ponaša, a valjanost je odnos između onog što se pretpostavlja da bi instrument (test) trebalo da meri i onog što on stvarno meri. Postoje tri standardna načina za ocenu pouzdanosti: interna konzistentnost, međuposmatračka pouzdanost i test-retest pouzdanost. Tri glavna tipa valjanosti su sadržinska, konstruktivna i kriterijumska. Uobičajeno se pouzdanost i valjanost ocenjuje primenom nekog od koeficijenata u studijama preseka, najčešće jedanput. U longitudinalnim studijama eliminiše se varijabilnost koja je posledica interindividualnih razlika pa se povećava verovatnoća za uočavanje značajne razlike nivoa faktora. Ponovljenim ocenama pouzdanosti i valjanosti dobiće se informacije o stabilnosti tih ocena u vremenu, a samim tim i pouzdanost i valjanost skorova dobijenih primenom skala.

Cilj: Ciljevi ovog istraživanja bili su: (1) Uraditi sistematski pregled literature i meta-analičko objedinjavanje ocena pouzdanosti i valjanosti psihijatrijskih skala saopštenih u različitim vremenima tokom longitudinalnih studija. (2) Proceniti promene ocena pouzdanosti i valjanosti psihijatrijskih skala saopštenih u različitim vremenima tokom longitudinalnih studija. (3) Generisati moderatore / prediktore prethodno navedenih ocena i njihovih promena.

Materijal i metode: Postavljeni cilj ovog istraživanja realizovan je u studiji slučaja sistematskog pregleda i meta-analize ocena pouzdanosti i valjanosti Hamiltonove skale za procenu depresivnosti (HAMD) i Montgomeri–Ašbergove skale za procenu

depresivnosti (MADRS). Ove dve skale su odabrane jer su najkorišćeniji psihijatrijski instrumenti koji se koriste od strane ocenjivača. Jedinice posmatranja su literaturne jedinice (publikovani radovi) koje su dobijene sistematskim pregledom literature u cilju identifikacije publikovanih istraživanja u kojima su saopštene ocene o pouzdanosti i/ili valjanosti skala HAMD i MADRS. Pretraživanje literature sprovedeno je u dve faze. U prvoj fazi pretražene su dve bibliografske (indeksno-abstraktne) baze podataka: MEDLINE i PsycINFO. Vremenski period obuhvaćen pretraživanjem proteže se od početka primene HAMD (1960. godine) i MADRS (1979. godine) skala, do kraja 2013 godine. Pretraživanjem baza podataka koje sadrže kompletan tekst objavljenih studija u elektronskoj formi (HINARI, EBSCO, ProQuest i Science Direct) identifikovane su i locirane dodatne studija za skale HAMD i MADRS. Rezultati pretraživanja su spojeni, a zatim su uklonjeni duplikati i lažno pozitivni nalazi. Ukupan broj prikupljenih studija u punom tekstu u kojima je korišćena HAMD skala iznosio 6590, a u kojima je korišćena MADRS skala iznosi 2051. U drugoj fazi izvršeno je pretraživanje ocena pouzdanosti i valjanosti u prikupljenim radovima u celini. Primarne studije pregledane su od strane dva nezavisna istraživača koji su u unapred pripremljen formular ekstrahovali podatke koji se odnose na ocene pouzdanosti i valjanosti HAMD i MADRS skala, kao i podatke koji se potencijalno mogu upotrebiti u analizi varijabli moderatora. Sistematski pregled i meta-analiza biće izvedeni za sve ocene pouzdanosti i valjanosti saopštene u dva vremena u longitudinalnim studijama. Veličina efekta u slučaju ocena pouzdanosti i koeficijenta korelacije između skala biće njihova zed transformisana vrednost. U cilju otkrivanja potencijalnih varijabli moderatora biće izvedena meta-regresija. Meta-analitičkim metodom robusne ocene varijanse modelovane su promene koeficijenta pouzdanosti i konkurentne valjanosti u vremenu, kao i povezanost tih promena sa potencijalnim prediktorima, dok su metodom recipročne varijanse objedinjeni koeficijenti pouzdanosti i konkurentne valjanosti u pojedinačnim vremenima.

Rezultati: Od 6590 prikupljenih i pregledanih studija za HAMD skalu 98 (1,5%) su identifikovane kao studije u kojima su saopšteni podaci o pouzdanosti ili valjanosti. Od 2051 prikupljenih i pregledanih studija za MADRS skalu 38 (1,9%) je identifikovano kao studije u kojima su saopšteni podaci o pouzdanosti ili valjanosti. Vrednosti Alfa koeficijenta za HAMD skalu više su na poslednjoj proceni kod 22 od 24 uključene studije. Postoje statistički značajne promene u vremenu zed transformisanih vrednosti alfa

koeficijenata za HAMD skalu ($p < 0,001$). Povećanje vrednosti alfa koeficijenta povezano je sa smanjenjem skora depresivnosti i povećanjem njegovog varijabiliteta. Vrednost zed transformisanih Loevingerinih koeficijenta za HAMD skalu raste u vremenu i taj porast je statistički značajan ($p = 0,002$). Nisu dobijene statistički značajne promene u vremenu za vrednosti zed transformisanih alfa i Loevingerinih koeficijenata za MADRS skalu, što je verovatno posledica malog uzorka. Postoji statistički značajan porast u vremenu zed transformisanih vrednosti intraklasnog koeficijenta korelacije za HAMD skalu ($p = 0,025$). U poređenju sa početnim merenjima, vrednosti intraklasnog koeficijenta korelacije za MADRS skalu više su na poslednjoj proceni. U poređenju sa početnim merenjima, vrednosti zed transformisanih koeficijenata test-retest pouzdanosti za HAMD skalu više su na poslednjoj proceni. Korelacije između skala za procenu depresivnosti povećavaju se tokom vremena kod 52 od 61 studije uključene u meta-analize. Statistički značajne promene u koeficijentima korelacije nađene su između skala HAMD i BDI ($p < 0,001$) i za korelacije između skala HAMD i MADRS ($p < 0,001$). Povećanje korelacija između skala povezano je sa smanjenjem skora depresivnosti i povećanjem njegovog varijabiliteta. Mali broj podataka nije mogao dovesti do zaključka o promenama dijagnostičke i faktorske valjanosti, kako za HAMD tako i za MADRS skalu.

Zaključci: Zaključci koji se odnose na meta-analitički metod ispitivanja promena pouzdanosti i valjanosti psihijatrijskih skala u vremenu su: (1) Postojećim meta-analitičkim metodama adekvatno se mogu oceniti pokazatelji pouzdanosti i valjanosti psihijatrijskih instrumenata u longitudinalnim studijama. (2) Meta-analiza ocena pouzdanosti i valjanosti u longitudinalnim studijama pruža objektivnu osnovu za generisanje preporuka njihove prihvatljive vrednosti koje su specifične za instrument/oblast i vreme primene. (3) Niska učestalost saopštavanja podataka o pouzdanosti i valjanosti primenjenih psihijatrijskih instrumenata predstavlja glavni problem za efektivno korišćenje meta-analitičkog metoda u istraživanju promena merne adekvatnosti u longitudinalnim studijama. Zaključci koji se odnose na meta-analizu promena ocena pouzdanosti i valjanosti za Hamiltonovu skalu za procenu depresivnosti i Montgomeri-Ašbergovu skalu za procenu depresivnosti: (1) Pouzdanost Hamiltonove skale za procenu depresivnosti, interna konzistenost, iskazana u vidu Cronbachovih alfa koeficijenata i Loevingerinih koeficijenata, kao i međuposmatračka pouzdanost, iskazana u vidu intraklasnih koeficijenata korelacije, pokazuje povećanje u vremenu u

longitudinalnim studijama. (2) Koeficijenti konkurentne valjanosti Hamiltonove skale za procenu depresivnosti i Montgomeri-Ašbergove skale za procenu depresivnosti iskazani u vidu koeficijenata korelacija, kao i koeficijenti korelacije sa Bekovom skalom za procenu depresivnosti pokazuju trend povećanja u longitudinalnim studijama. Njihove vrednosti su za oko trećinu veće na kraju studija u odnosu na početak. Iz navedenog se može zaključiti da je mali broj publikovanih podataka u ovom tipu istraživanja glavna prepreka za dostizanje opšteg zaključka o promenama pouzdanosti i valjanosti psihijatrijskih skala u vremenu u longitudinalnim studijama.

Ključne reči: pouzdanost, valjanost, skale za procenu depresivnosti, longitudinalne studije, meta-analiza

Naučna oblast: Medicina

Uža naučna oblast: Medicinska statistika

SYSTEMATIC REVIEW AND META-ANALYSIS OF RELIABILITY AND VALIDITY OF PSYCHIATRIC SCALES IN LONGITUDINAL STUDIES

Bukumirić M. Zoran

Abstract:

Introduction: Rating scales in psychiatry make it possible to detect the presence or lack of symptoms of the psychiatric disorders, to express the intensity of symptoms in numbers and to use it for clinical or research purposes. Thus, an important aspect of these scales are their metric characteristics, their measurement adequacy what includes reliability and validity. While reliability indicates the extent to which the method of measurement (test) behaves consistently and steadily, the validity is the ratio between what the instrument (test) is supposed to measure and what it really measures. There are three standard ways of evaluating reliability: internal consistency reliability, inter-rater reliability and test-retest reliability. Three main types of validity are content validity, construct validity and criterion-related validity. Commonly reliability and validity are evaluated by using one of the coefficients in the cross-sectional studies, usually once. In longitudinal studies, variability which is a consequence of interindividual differences is eliminated and thus the chance of spotting significant difference in factor levels is increased. By repeated assessments of reliability and validity information on score stability over time will be obtained, and consequently the reliability and validity of the scores obtained by used scales.

Aims of study: The aims of this study were to: (1) Conduct a systematic review and meta-analysis of the reliability and validity scores of psychiatric scales used in longitudinal studies at different points in time. (2) Evaluate changes in the reliability and validity scores in longitudinal studies at different points in time. (3) Generate moderators/predictors of mentioned scores and their changes.

Material and methods: The aim of research was carried out through case study of systematic review and meta-analysis of the reliability and validity of Hamilton Depression Rating Scale (HAMD) and Montgomery-Asberg Depression Rating Scale (MADRS). These two rating scales were chosen because they are the psychiatric

instruments most widely used by the assessors. The units of analysis were an individual studies (published papers) obtained by systematic review of the literature in order to identify published research in which the assessment of the reliability and/or validity of HAMD and MADRS rating scales is reported. Literature search was performed in two stages. In Stage I, we searched two bibliographic (index-abstract) databases: MEDLINE and PsycINFO. Period covered by search extends from the beginning of implementation of HAMD (1960) and MADRS (1979) rating scales to the end of the year 2013. By searching databases containing the entire text of published studies in electronic format (HINARI, EBSCO, ProQuest and Science Direct) additional studies for HAMD and MADRS rating scales have been identified and located. After results were pooled, false positive records were removed and duplicates resolved. The total number of collected articles in full text was 6590 for HAMD rating scale and 2051 for HAMD rating scale.

In Stage II, pooled full-text papers were searched for reliability and validity scores. Primary studies were reviewed independently by two researchers who extracted information related to the assessment of the reliability and validity of HAMD and MADRS rating scales in pre-prepared form, as well as data that can potentially be used in the analysis of moderator variables. A systematic review and meta-analysis was performed for all ratings of reliability and validity reported in two time points in longitudinal studies. Effect sizes of reliability scores and correlation coefficients between scales was their z-transformed value. In order to detect potential moderator variables meta-regression was applied. Relationship between changes in coefficients of reliability and concurrent validity in time with potential predictors were modelled using robust variance meta-regression, while the pooled coefficients of reliability and concurrent validity in individual points of time were obtained by the inverse variance meta-analytic method.

Results: Out of 6590 collected and reviewed studies for HAMD rating scale 98 (1.5%) were identified as studies in which data on reliability and validity were given. At the same time out of 2051 collected and reviewed studies for MADRS rating scale 38 (1.9%) were identified as studies in which data on reliability or validity were provided. Alpha coefficient values for HAMD rating scale were higher at last evaluation in 22 out of 24 included studies. Significant changes in time of z-transformed alpha coefficient values for HAMD rating scale ($p < 0.001$) were found. Increase in the alpha coefficient value is

associated with a decrease in depression score and increase in its variability. Value of z-transformed Loevinger coefficients for HAMD rating scale is increased in time and the increase is statistically significant ($p=0.002$). Changes in z-transformed alpha and Loevinger coefficients values in time for MADRS rating scale were not statistically significant, most probably due to small sample size. Increase of z-transformed values of intraclass correlation coefficient in time for HAMD rating scale is statistically significant ($p=0.025$). In comparison to initial measurements, values of intraclass correlation coefficient for MADRS rating scale are higher at last evaluation. In comparison to initial measurements, values of z-transformed coefficient for test-retest reliability for HAMD rating scale are higher at last evaluation. Correlation coefficients between depression scales increase over time in 52 out of 61 studies included in meta-analysis. Statistically significant changes in correlation coefficients are found between HAMD and BDI rating scales ($p<0.001$), and between HAMD and MADRS rating scales ($p<0.001$). Increase in correlation between rating scales is associated with a decrease in depression score and increase in its variability. Due to small sample size conclusion on changes of diagnostic and factorial validity for both HAMD and MADRS rating scales could not be drawn.

Conclusions: Conclusions related to meta-analytic method in assessment of changes in reliability and validity of psychiatric rating scales over time are: (1) Existing meta-analytic methods can adequately be used in assessment of reliability and validity indicators of psychiatric instruments in longitudinal studies. (2) Meta-analysis evaluation of reliability and validity in longitudinal studies provides fair basis for creation of recommendation on their acceptable value which are specific for instrument/field and time of their implementation. (3) Low frequency of reporting data on reliability and validity of applied psychiatric instruments is a major problem for the effective use of meta-analytic methods in the study of changes in measurement adequacy in longitudinal studies. Conclusions related to meta-analysis of changes in reliability and validity assessments for Hamilton rating scale for depression and Montgomery-Asberg depression rating scale: (1) The reliability of the Hamilton rating scale for depression, internal consistency, expressed as Cronbach's alpha coefficients and Loevinger coefficients as well as and inter-rater reliability, expressed as intraclass correlation coefficients, indicates an increase over time in longitudinal studies. (2) Concurrent validity coefficients for Hamilton rating scale for depression and Montgomery-Asberg depression rating scale

expressed as correlation coefficients, as well as correlation coefficients with Beck depression inventory show rising trend in longitudinal studies. Their values are about a third higher at the end of the study compared to beginning. From the above mentioned it can be concluded that the small number of published data in this type of research is the main obstacle in reaching a general conclusion about the changes in the reliability and validity of psychiatric rating scales over time in longitudinal studies.

Key words: reliability, validity, depression rating scales, longitudinal studies, meta-analysis

Scientific area: Medicine

Special topics: Medical statistics

Sadržaj

1. Uvod	1
1.1. Psihijatrijske skale	1
1.2. Merenje u psihijatriji upotrebom skala	2
1.3. Pouzdanost skala u psihijatriji	3
1.3.1. Interna konzistentnost	4
1.3.1.1. Alfa koeficijent	4
1.3.1.2. Kuder-Richardsonova formula 20	5
1.3.1.3. Loevingerin koeficijent	5
1.3.1.4. Pouzdanost dve polovine	5
1.3.2. Međuposmatračka pouzdanost	6
1.3.3. Test-retest pouzdanost	7
1.3.4. Pouzdanost paralelnih formi	7
1.4. Valjanost skala u psihijatriji	8
1.4.1. Sadržinska valjanost	8
1.4.2. Kriterijumska valjanost	8
1.4.3. Konstruktivna valjanost	9
1.5. Odnos pouzdanosti i valjanosti	9
1.6. Meta-analiza	10
1.6.1. Sistematski pregled i meta-analiza	10
1.6.2. Veličina efekta	11
1.6.3. Metode meta-analize	12
1.6.4. Modeli meta-analize	13
1.6.5. Procena publikacione pristrasnosti	14
1.6.6. Heterogenost	15
1.6.7. Analiza moderatora	15
1.6.8. Analiza senzitivnosti	16

1.7. Meta-analiza ocena pouzdanosti i valjanosti	16
1.8. Psihometrijske osobine Hamiltonove skale za procenu depresivnosti	18
1.9. Psihometrijske osobine Montgomeri-Ašbergove skale za procenu depresivnosti 19	
1.10. Psihometrijske osobine Bekove skale za procenu depresivnosti.....	20
1.11. Longitudinalne studije	21
2. Ciljevi istraživanja.....	22
3. Materijal i metod istraživanja.....	24
3.1. Pretraživanje literature, selekcija studija i ekstrakcija podataka	24
3.2. Kriterijumi za uključivanje, neuključivanje i isključivane studija iz meta- analize.....	28
3.3. Ekstrakcija podataka.....	29
3.4. Analiza podataka	33
3.4.1. Korekcija i transformacija podataka.....	35
3.4.1.1. Spearman-Brownova korekcija	35
3.4.1.2. Fisherova z-transformacija	35
3.4.2. Robusna ocena varijanse u meta-regresiji sa veličinom efekta kao zavisnom varijablom.....	36
3.4.3. Meta-analitički metod Hedgesa i Olkina	38
3.4.3.1. Model sa fiksnim efektima	38
3.4.3.2. Model sa slučajnim efektima.....	39
3.4.4. Analiza podataka o dijagnostičkoj valjanosti	39
3.4.4.1. Rekonstrukcija dijagnostičke tabele	39
3.4.4.2. Objedinjavanje ocena dijagnostičkih performansi	40
3.4.4.3. Testiranje homogenosti ocena dijagnostičkih performansi	40
3.4.5. Procena publikacione pristrasnosti	41
3.4.6. Heterogenost.....	41
3.4.7. Analiza moderatora.....	42
3.4.8. Analiza senzitivnosti	43
3.4.9. Softver korišćen za analizu podataka	43

4. Rezultati.....	44
4.1. Objedinjavanje pouzdanosti u vremenu	46
4.1.1. Interna konzistentnost za HAMD skalu	46
4.1.1.1. Analiza alfa koeficijenta za HAMD skalu.....	46
4.1.1.2. Analiza Loevingerinog koeficijenta za HAMD skalu.....	54
4.1.1.3. Pouzdanost dve polovine.....	61
4.1.2. Interna konzistentnost za MADRS skalu.....	62
4.1.2.1. Analiza alfa koeficijenta za MADRS skalu.....	62
4.1.2.2. Analiza Loevingerinog koeficijenta za MADRS skalu.....	68
4.1.3. Međuposmatračka pouzdanost	70
4.1.3.1. Analiza ocena međuposmatračke pouzdanosti ukupnih skorova za HAMD skalu	70
4.1.3.1.1. Analiza intraklasnog koeficijenta korelacije za HAMD skalu	71
4.1.3.1.2. Analiza Kappa koeficijenta za HAMD skalu	78
4.1.3.2. Analiza ocena međuposmatračke pouzdanosti ukupnih skorova za MADRS skalu	79
4.1.3.2.1. Analiza intraklasnog koeficijenta korelacije za MADRS skalu	80
4.1.3.2.2. Analiza kappa koeficijenta za MADRS skalu	83
4.1.4. Test-retest pouzdanost	85
4.1.4.1. Analiza ocena test-retest pouzdanosti za HAMD skalu	85
4.2. Objedinjavanje valjanosti u vremenu	88
4.2.1. Analiza ocena korelacija.....	88
4.2.2. Analiza dijagnostičke valjanosti.....	105
4.2.2.1. Analiza dijagnostičke valjanosti za HAMD skalu.....	105
4.2.2.2. Analiza dijagnostičke valjanosti za MADRS skalu.....	110
4.2.3. Analiza faktorske valjanosti	113
4.2.3.1. Analiza faktorske valjanosti za HAMD skalu	113
4.2.3.2. Analiza faktorske valjanosti za MADRS skalu	115
5. Diskusija.....	116
6. Zaključci.....	132

7. Literatura	134
Prilozi	156
Prilog 1 – R komande korišćene u analizi podataka.....	157
Prilog 2 – Studije u kojima su saopštene vrednosti alfa koeficijenta	163
Prilog 3 – Studije u kojima su saopštene vrednosti Loevingerinog koeficijenta.....	169
Prilog 4 – Studije u kojima su saopštene ocene međuposmatračke pouzdanosti ukupnih skorova	172
Prilog 5 – Studije iz kojih su, sekundarnom analizom, izračunate ocene test-retest pouzdanosti u ponovljenim merenjima	180
Prilog 6 – Studije u kojima su saopštene korelacije između skala HAMD, MADRS i BDI.....	182
Prilog 7 – Studije u kojima su saopštene ocene dijagnostičke valjanosti.....	192
Prilog 8 – Studije u kojima su saopštene ocene faktorske valjanosti.....	195
Spisak skraćenica	199

1. Uvod

1.1. Psihijatrijske skale

Pouzdanost, zajedno sa valjanošću, je veoma važna osobina bilo kog psihijatrijskog instrumenta. Dobiti valjane, odnosno tačne podatke, od primarnog je značaja za svakog kliničara tokom praćenja pacijenata vezano za ishod terapije ili progresiju psihijatrijskih simptoma, u kliničkoj praksi i istraživanjima (Blacker, 2009; Möller, 2009).

Depresija je privatno subjektivno iskustvo i nije je moguće direktno meriti. Procena depresivnog raspoloženja u velikoj meri zavisi od iskaza osobe koja to raspoloženje doživljava. U poslednjih nekoliko decenija, mnogi instrumenti su razvijeni za procenu depresivnog raspoloženja i depresije. Među njima, Hamiltonova skala za procenu depresivnosti (*Hamilton Depression Rating Scale* (engl.) – HAMD) (Hamilton, 1960) i Montgomery-Ašbergova skala za procenu depresivnosti (*Montgomery-Åsberg Depression Rating Scale* (engl.) – MADRS) (Montgomery i Asberg, 1979) su skale procene, dok je Bekova skala za procenu depresivnosti (*Beck Depression Inventory* (engl.) – BDI) (Beck i sar., 1961) skala samoprocene.

Korišćenjem ovih skala smanjuje se subjektivnost u proceni depresije. Psihijatrijske skale omogućavaju da se registruje prisustvo ili odsustvo simptoma psihijatrijskih oboljenja, da se njihov intenzitet izrazi brojevima i upotrebi u kliničke ili istraživačke svrhe. Skale nisu zamena, već su dopuna kliničkoj proceni i dijagnozi depresije (Carneiro i sar., 2015; Schrag i sar., 2007). Pouzdani i valjani instrumenti su neophodni da potvrde dijagnozu depresije, smanje pristrasnost izazvanu fizičkim simptomima, procene napredak tokom tretmana i evaluiraju ishod (Cusin i sar., 2009). Često se dešava da istraživači istovremeno koriste različite instrumente za sveobuhvatnu procenu depresije. Neki autori predlažu istovremenu upotrebu skala procene i skala samoprocene, da bi se pokrili različiti domeni poput depresivnih simptoma i funkcionisanja (Lin i sar., 2014; Möller, 2000; Uher i sar., 2008).

Važan aspekt skala u psihijatriji jesu njihove metrijske osobine, odnosno njihova merna adekvatnost što obuhvata pouzdanost i valjanost. Pouzdanost označava do koje mere se metoda merenja (test) konzistentno ili postojano ponaša, a valjanost je odnos između onog što se pretpostavlja da bi instrument (test) trebalo da meri i onog što on stvarno meri (Erić-Marinković i sar., 2001).

Podaci o pouzdanosti i valjanosti skala variraju od studije do studije. Meta-analizom ocena pouzdanosti i valjanosti dobija se odgovor na pitanje koje su to tipične vrednosti koeficijenata pouzdanosti i valjanosti, kakav je njihov varijabilitet od studije do studije, i koji su to faktori kojima se može objasniti varijabilitet ovih koeficijenata. U ovom istraživanju biće prikazane meta-analiitičke metode za dobijanje objedinjene ocene o pouzdanosti i valjanosti psihijatrijskih skala u longitudinalnim studijama.

1.2. Merenje u psihijatriji upotrebom skala

Prema načinu primene skale se mogu podeliti na skale procene i skale samoprocene. Skale procene primenjuju ocenjivači. Ocenjivači mogu biti: psihijatri i lekari opšte, istraživači, psiholozi i socijalni radnici, medicinski tehničari. Skale samoprocene popunjava sam ispitanik.

Pouzdanost i valjanost se odnose na tumačenje skorova iz psihometrijskih instrumenata (Cook i Beckman, 2006). Pouzdanost se, u najširem smislu, može definisati kao merenje oslobođeno greške. Valjanost se, u najširem smislu, odnosi na stepen sa kojim instrument meri ispitivani konstrukt (Cohen i Swerdlik, 2009). Neophodan, ali ne i dovoljan uslov valjanosti merenja je da je ono i pouzdano (Baer i Blais, 2009; Moss, 1994).

Početak tradicionalnog pristupa pouzdanosti vezuje se za radove Spearmana (Spearman, 2010; Spearman, 1910). Njegov model *pravog skora i skora greške* je ostao najuticajni model u psihološkim istraživanjima. Kasnije su tu teoriju razvili Lord i Novick i poznata je kao *klasična teorija testa* ili *teorija pravog skora* (Lord, 1959; Novick, 1966). Po ovoj teoriji opservirani skor jednog ispitanika na skali sadrži dve

komponente. Jedna komponenta predstavlja pravu (tačnu) meru varijable, dok drugu komponentu čini skor greške u merenju.

$$\begin{array}{rcc} X & = & T & + & E \\ \text{opservirani skor} & & \text{pravi skor} & & \text{greška merenja} \end{array}$$

U ovoj jednačini, poznata je samo vrednost dobijenog skora, dok su druge dve veličine, pravi skor i skor greške, nepoznate. Skor greške je odstupanje dobijenog skora od pravog skora, i nastaje kao rezultat greške merenja čiji je izvor u nepouzdanosti skale.

Varijansa opserviranog skora (s_o^2) se takođe može rastaviti na dve komponente: na varijansu pravog skora (s_t^2) i varijansu skora greške (s_e^2):

$$s_o^2 = s_t^2 + s_e^2$$

Koeficijent pouzdanosti se teorijski definiše kao odnos varijanse pravog skora i varijanse dobijenog skora:

$$r_{xx} = \frac{s_t^2}{s_o^2}$$

Iz jednačine se vidi da koeficijent pouzdanosti ukazuje na to u kojoj meri dobijeni varijabilitet na testu (s_o^2) sadrži pravi varijabilitet (s_t^2).

1.3. Pouzdanost skala u psihijatriji

Pouzdanost skorova skala u psihijatriji direktno odgovara stepenu prisustva ili odsustva greške u bihevioralnoj meri. Pouzdanost se može posmatrati kao preciznost instrumenta. U zavisnosti šta se posmatra kao glavni izvor varijanse skora greške, pouzdanost se može oceniti putem ocene interne konzistentnosti, međuposmatračke pouzdanosti, test-retest pouzdanosti i pouzdanosti paralelnih formi. Zavisno od skale merenja skorova (kategorijalna ili kontinuirana), razlikuju se načini izračunavanja ovih ocena.

1.3.1. Interna konzistentnost

Interna konzistentnost je mera slaganja individualnih stavki (ajtema) na skali i označava stepen do kojeg su stavke međusobno korelisane. Ovom prilikom, svaka stavka se posmatra kao poseban instrument za merenje ispitivanog konstrukta, tako da koherentnost između njih sugeriše da svi ajtemi mere istu stvar (a takođe i sve stavke zajedno kao celina). Interna konzistencija skale zavisi od interne konzistencije konstrukta kojeg težimo da tom skalom izmerimo, i veća je kod unidimenzionalnih konstrukata nego kod onih sa dva ili više nezavisnih domena.

Interna konzistentnost se može oceniti: *alfa koeficijentom* (Cronbach, 1951), *Kuder-Richardsonovom formulom 20* (KR-20) (Kuder i Richardson, 1937), *Loevingerinim koeficijentom* (Loevinger, 1948), *pouzdanošću dve polovine* sa prilagođavanjem Spearman-Brown ili Guttmanovom formulom (Kaplan i Saccuzzo, 2013).

1.3.1.1. Alfa koeficijent

Alfa koeficijent je najopštiji način za nalaženje ocene pouzdanosti kroz internu konzistenciju. On ocenjuje pouzdanost skora testa koristeći informacije odnosa između ajtema testa. Prema imenu autora često se naziva i Cronbach alfa. To je, možda, najviše upotrebljavan koeficijent pouzdanosti (Webb i sar., 2006).

Formula za izračunavanje alfa koeficijenta:

$$r = \alpha = \left(\frac{n}{n-1} \right) \left(\frac{s^2 - \sum s_i^2}{s^2} \right)$$

gde je n broj ajtema u skali, s_i^2 varijansa ajtema, a s^2 je varijansa skale.

1.3.1.2. Kuder-Richardsonova formula 20

Izračunava pouzdanost testa u kome su ajtemi dihotomni, skorovi 0 i 1 (najčešće za ispravno i pogrešno).

$$KR_{20} = r = \left(\frac{n}{n-1} \right) \left(\frac{s^2 - \sum pq}{s^2} \right)$$

gde je KR_{20} ocena pouzdanosti (r), n je broj ajtema na testu, s^2 je varijansa ukupnog skora na testu, p je proporcija ispitanika koji su odgovorili ispravno za svaki ajtem, q je proporcija ispitanika koji nisu odgovorili ispravno za svaki ajtem.

1.3.1.3. Loevingerin koeficijent

Loevingerin koeficijent homogenosti koristi se za procenu hijerarhijske strukture ajtema unutar skale (Bjerg Bendsen i sar., 1997; Loevinger, 1948).

$$H_t = \frac{V_x - V_{het}}{V_{hom} - V_{het}}$$

gde je H_t homogenost testa, V_x je varijansa testa, V_{het} je varijansa hipotetički perfektno heterogenog testa sa istom distribucijom ajtema, V_{hom} je varijansa hipotetički perfektno homogenog testa sa istom distribucijom ajtema.

1.3.1.4. Pouzdanost dve polovine

Alternativni indeks interne konzistencije je *pouzdanost dve polovine*¹. Izračunava se u tri koraka: (1) skala se proizvoljno podeli na dva dela i izračuna ukupan skor za dve polovine, (2) izračuna se korelacija ukupnih skorova dve polovine, (3) korelacija se prilagodi preko Spearman-Brown ili Guttmanove formule (Cohen i Swerdlik, 2009).

¹ *Split-half reliability* (engl.)

Spearman-Brown model se bazira na korelaciji između delova skale podeljene na dve polovine, izračunava se preko formule:

$$r_{tt} = \frac{2r_{hh}}{(1 + r_{hh})}$$

gde je r_{tt} – pouzdanost ukupnog testa dobijena iz pouzdanosti jedne od polovina (koeficijent pouzdanosti kompletnog testa), a r_{hh} – samo-korelacija polovine testa (koeficijent pouzdanosti polovine testa).

Guttmanova formula je drugi način procene pouzdanosti kompletne skale (Guttman, 1945). Ne pretpostavlja da su oba dela testa podjednako pouzdana i da imaju istu varijansu, pa je dobijeni koeficijent pouzdanosti manji. Izračunava se na osnovu formule:

$$r_{tt} = 1 - \frac{SD_d^2}{SD_x^2}$$

gde je: SD_d^2 varijansa razlika između pojedinačnih skorova dve polovine testa, a SD_x^2 je varijansa ukupnog skora.

1.3.2. Međuposmatračka pouzdanost

Međuposmatračka pouzdanost se koristi za procenu stepena slaganja dva ili više ocenjivača koji koriste iste informacije o istom objektu. Kod skala samoprocene ne izračunava se međuposmatračka pouzdanost.

Postoji više postupaka za izračunavanje koeficijenata međuposmatračke pouzdanosti. Najčešće se koriste intraklasni koeficijent korelacije² (*ICC*), Pearsonov koeficijent korelacije, kappa koeficijent, Spearmanov koeficijent korelacije i Kendallov koeficijent konkordancije.

² *Intraclass correlation coefficient* (engl.)

ICC je mera proporcije varijanse koja se može pripisati objektima merenja odnosno proporcija varijanse unutar predmeta ocenjivanja (Bartko, 1966; Hallgren, 2012). Formula za izračunavanje intraklasnog koeficijenta korelacije je:

$$ICC = \frac{\sigma^2(b)}{\sigma^2(b) + \sigma^2(w)}$$

gde je $\sigma^2(w)$ pulovana varijansa unutar ispitanika, a $\sigma^2(b)$ je varijansa između predmeta merenja. Imenilac $\sigma^2(w) + \sigma^2(b)$ predstavlja ukupnu varijansu svih ocena bez obzira na predmet merenja.

1.3.3. Test-retest pouzdanost

Test-retest pouzdanost je ocena stabilnosti skale u vremenu, odnosno mera slaganja rezultata koji se dobijaju ponovljenim merenjem na istim objektima pod uslovom da nije došlo do promene stanja tih objekata. Ocenjuje se primenom: intraklasnog koeficijenta korelacije, Pearsonovog koeficijenta linearne korelacije ili Spearmanovog koeficijenta rang korelacije.

Ova vrsta pouzdanosti je posebno značajna ako se skala koristi za ocenu promene stanja ispitanika u toku tretmana. Vremenski interval 2-3 nedelje može biti adekvatan za procenu test-retest pouzdanosti instrumenata za ocenu depresivnosti (Baer i Blais, 2009). Kod skala sa niskom test-retest pouzdanošću nemoguće je odrediti da li je izmerena promena stvarna ili predstavlja slučajnu grešku skale (Cook i Beckman, 2006).

1.3.4. Pouzdanost paralelnih formi

Podrazumeva primenu različitih verzija instrumenta na istim ispitanicima u isto ili različito vreme. Korelacija između dve paralelne forme je ocena pouzdanosti (Cohen i Swerdlik, 2009).

Pouzdanost paralelnih formi je slična pouzdanosti dve polovine. Glavna razlika je u tome što su kod paralelne forme obe forme testa konstruisane da se mogu koristiti nezavisno jedna od druge i smatraju se ekvivalentnim merama (Cook i Beckman, 2006).

1.4. Valjanost skala u psihijatriji

Valjanost se odnosi na stepen slaganja između skora na testu i konstrukta (npr. beznađe) ili dijagnoze (npr. depresija) za koje je dizajnirana da meri (Baer i Blais, 2009). Pouzdanost je empirijska i može se izračunati odmah nakon primene skale. Valjanost je delom teorijska jer mnogi konstrukti mereni u psihijatriji nemaju apsolutnu istinu odnosno jasno dokazan biološki supstrat, pa je često potrebno nekoliko godina da bi bila ocenjena tačno. Valjanost mora biti potvrđena sa više kriterijuma, empirijski dokazi treba da pokažu da instrument meri ono za šta je namenjen da meri

Valjanost se najčešće deli na tri glavna tipa: sadržinsku, kriterijumsku i konstruktivnu valjanost (Cohen i Swerdlik, 2009; Guion, 1980; Kaplan i Saccuzzo, 2013). Postoji više naziva za različite aspekte valjanosti ali se svi mogu svrstati u navedena tri tipa (Kaplan i Saccuzzo, 2013).

1.4.1. Sadržinska valjanost

Sadržinska valjanost ukazuje na to da li instrument u potpunosti predstavlja konstrukt koji meri (Blacker, 2009). *Prividna valjanost* se odnosi na to da li svaki od ajtema meri konstrukt od interesa.

1.4.2. Kriterijumska valjanost

Kriterijumska valjanost je empirijski zasnovana i odnosi se na povezanost skale sa nekim spoljnim kriterijumom. Obično se meri korelacijom sa drugim skalama, čija je valjanost u merenju istog konstrukta već ustanovljena (McDowell, 2006; Portney i Watkins, 2009; Streiner i Norman, 2014). Postoje dva tipa kriterijumske valjanosti: prediktivna i konkurentna. *Prediktivna valjanost je korelacija* sa kriterijumom koji nije poznat u vreme primene skale, već će biti poznat nekoliko dana ili nekoliko godina kasnije. *Konkurentna valjanost* je korelacija nove skale sa spoljnim kriterijumom koji se primenjuje istovremeno. Dva tipa konkurentne valjanosti su konvergentna i diskriminaciona valjanost. *Konvergentna valjanost* je korelacija sa nekom varijablom, sa kojom bi inače skala trebalo (teorijski) da bude povezana, ili korelacija sa drugom skalom

koja meri isti ili sličan konstrukt. *Dijagnostička valjanost* se može svrstati u konvergentnu valjanost gde se slaganje podataka ocenjuje u odnosu na zlatni standard. *Diskriminaciona valjanost* se može svrstati i u kriterijumsku i konstruktivnu valjanost, a ispituje sposobnost instrumenta da diskriminiše populacije za koje se očekuje da se razlikuju na posmatranom konstruktu (Myers i Winters, 2002).

1.4.3. Konstruktivna valjanost

Konstruktivna valjanost se primenjuje kada nema adekvatnog zlatnog standarda, postojećih instrumenata, kriterijuma tačnosti ili drugih raspoloživih podataka za validaciju. Na primer za varijable bol, kvalitet života ili sreća ne postoji zlatni standard, pa se konstruktivna valjanost utvrđuje poređenjem sa spoljnim validatorima odnosno pojavama koje imaju jasne veze sa konstruktom od interesa ali nisu merene direktno instrumentom. Dva specijalna slučaja konstruktivne valjanosti su diskriminaciona valjanost (opisana u okviru kriterijumske valjanosti) i senzitivnost na promene (Blacker, 2009; McDowell, 2006). *Senzitivnost na promene* je sposobnost instrumenta da pokaže očekivane promene kliničkog stanja u toku tretmana ili promene u progresivnoj bolesti. Uslov da instrument ima dobru valjanost ovog tipa je da ima dobru test-retest pouzdanost.

Poseban pristup konstruktivnoj valjanosti je *faktorska valjanost*, koja se ispituje statističkom analizom pod nazivom faktorska analiza (Myers i Winters, 2002). Faktorskom analizom se može utvrditi da li je skala jednodimenzionalna ili višedimenzionalna, odnosno da li meri jedinstven konstrukt ili ne. Pod pretpostavkom jednodimenzionalnosti skale idealno rešenje je da se u faktorskoj analizi dobije jedan faktor, a za skale koje se sastoje od podskala više faktora (Fajgelj, 2013; Myers i Winters, 2002).

1.5. Odnos pouzdanosti i valjanosti

Da bi instrument bio valjan neophodno je da bude pouzdan. Koncept valjanosti srodan je konceptu tačnosti, a koncept pouzdanosti srodan je konceptu preciznosti. Metaforično rečeno, neko ko uči streljaštvo mora prvo da nauči kako da pogodi centar mete, a zatim kako da to radi konzistentno, što je analogno valjanosti i pouzdanosti

merjenja. Konzistentnost (ili pouzdanost) merjenja može se predstaviti sa koliko blizu će uzastopni pogoci biti jedan drugom, bilo gde na meti. Valjanost predstavlja koliko blizu, u proseku, pogoci idu u centar mete (McDowell, 2006).

Kada je pouzdanost skale niska nastaje snižavanje koeficijenta valjanosti (atenuacija). Poboljšanje pouzdanosti povećava valjanost sve do određenog stepena kada visoka pouzdanost počinje da vodi snižavanju kriterijumske valjanosti (paradoks atenuacije) (Blacker, 2009; Fajgelj, 2013). Rešenje problema ponudio je Guilford (1954). a sastoji se u umerenosti, tj. da se u postupku konstrukcije skale biraju ajtemi čije su ajtem-total korelacije od 0,30 do 0,80, što povlači interkorelaciju ajtema od 0,10 do 0,60 (Guilford, 1954).

1.6. Meta-analiza

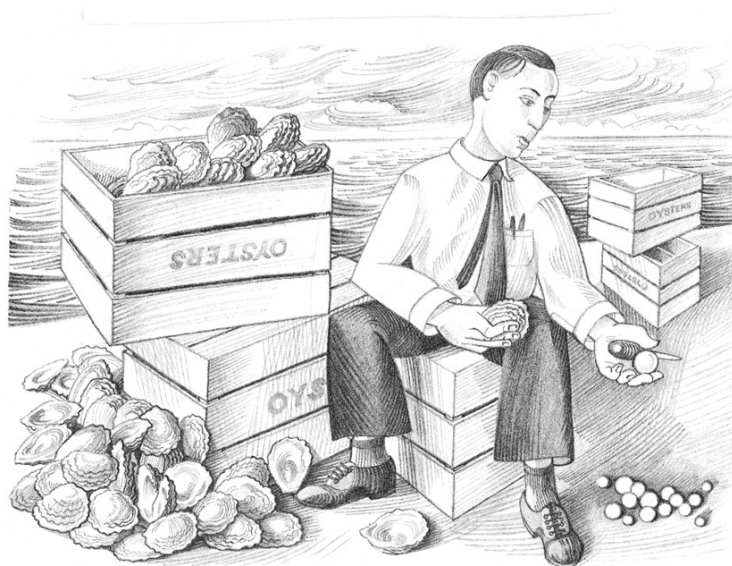
1.6.1. Sistematski pregled i meta-analiza

Sistematski pregled identifikuje i objedinjava dokaze koji se odnose na određeno istraživačko pitanje. Sistematski pregled može biti: (a) kvalitativan ili (b) kvantitativan (meta-analiza). Sistematski pregled se odnosi na ceo proces pronalaženja, selekcije, procena, sinteze i izveštavanja dokaza, a meta-analiza je specifična statistička tehnika kombinovanja podataka iz pojedinačnih studija (Borenstein i sar., 2011; Cooper i sar., 2009; Higgins i Green, 2008; Lipsey, 2011). Prema rečima Glassa: „Meta-analiza se odnosi na analizu analiza...statističku analizu velike zbirke rezultata analiza iz pojedinih istraživanja sa svrhom integrisanja rezultata“ (Glass, 1976).

U današnje vreme, u medicini, postoji sve veće interesovanje za meta-analizu. U 2000. godini na PubMedu indeksirano je 1064 meta-analiza, dok je taj broj 2014. godine iznosio 14519. Ovaj trend će se nastaviti zbog eksponencijalnog rasta objavljenih istraživanja i široke dostupnosti softvera i skripta za izvođenje meta-analize (Quintana, 2015).

Koraci tokom izvođenja sistematskog pregleda i meta-analize: (1) Formulacija istraživačkog pitanja, (2) Određivanje kriterijuma za uključivanje studija u analizu,

(3) Pretraživanje i lociranje relevantnih studija, (4) Utvrđivanje konačnog skupa studija, (5) Ekstrakcija podataka i kodiranje karakteristika studija, (6) Donošenje odluke o tome da li su podaci nezavisni ili zavisni, (7) Detekcija ekstremnih vrednosti, (8) Analiza podataka – kvalitativna ili kvantitativna (meta-analiza), (9) Interpretacija rezultata i donošenje zaključaka (Cooper Harris M., 1998; Littell i sar., 2008).



Slika 1. Metaforičan prikaz ekstrakcije podataka iz studija u meta-analizi (preuzeto iz Greenhalgh, 1997)

1.6.2. Veličina efekta

Meta-analiza počinje sumiranjem podataka iz svake studije radi izračunavanja veličine efekta studije. Veličina efekta je broj koji odražava veličinu odnosa između dve varijable. U prikupljenim studijama postoje 4 tipa informacija koje se mogu iskoristiti za veličinu efekta: (a) saopštena veličina efekta, (b) informacije koje se mogu iskoristiti za izračunavanje veličine efekta, (c) informacije da li je prilikom testiranja hipoteza dobijen statistički značajan odnos između nezavisne i zavisne varijable i smer tog odnosa, (d) informacija o smeru odnosa između nezavisnih i zavisnih varijabli (Cooper i sar., 2009).

Postoji više različitih metrika koje se mogu iskoristiti za merenje veličine efekta. Lipsey (2011) je predložio podelu veličine efekta na osnovu odnosa varijabli: (1) Odnos

dve kontinuirane varijable: podaci su dati u vidu koeficijenata korelacije. (2) Odnos jedne dihotomne i jedne kontinuirane varijable: podaci predstavljaju aritmetičke sredine, standardne devijacije i veličine uzoraka. (3) Odnos između dve dihotomne varijable: podaci predstavljaju frekvencije ili proporcije iz tabele kontingencije 2×2 . (a) razlika između dve verovatnoće, (b) odnos dve verovatnoće, (c) Fi koeficijent i (d) Odnos šansi.

Metrike veličine efekta mogu se konvertovati, kao što je mogućnost konverzije koeficijenta korelacije iz r u Fisherov z , i iz r u d (Schulze, 2007).

U meta-analizi se obično koriste težinski koeficijenti koji zavise od uzoračke tačnosti veličine efekta koja je dobijena iz studije. U tu svrhu se najčešće koriste veličina uzorka, ili neka veličina koja je funkcija veličine uzorka kao što je uzoračka varijansa veličine efekta.

1.6.3. Metode meta-analize

Postoji više tipova meta-analitičkih metoda. Bangert-Drowns (1986) opisuje pet metoda: (1) Meta-analitički metod Glassa (1976); (2) Bliska Glassovom metodu je “meta-analiza efekata studija”; (3) “Metod kombinovanja verovatnoća” (Rosenthal i Rubin, 1982a, 1982b; Rosenthal, 1978); (4) “Približno objedinjavanje podataka sa testiranjem homogenosti” (Hedges i Olkin, 1982, 1993; Hedges, 1981; Rosenthal i Rubin, 1982a, 1982b); (5) “Približno objedinjavanje podataka sa korekcijom uzoračke greške” (Hunter i sar., 1982; Schmidt i Hunter, 2014).

Johnson i sar. (1995) naglašavaju tri glavna meta-analitička metoda: metod Hedgesa i Olkina (1985), Rosental i Rubin metod (1978, 1988) i metod Hunter, Schmidt i Jackson (1982).

Prema Schulzeu i sar. (2007) dve dominantne metode meta-analize u psihologiji su metode prezentovane od strane Huntera i Schmidta, i od Hedgesa i Olkina. Pored toga postoji veliki broj poboljšanja i varijanti koje su slične ili identične ovim dvema metodama.

Novije metode meta-analize obuhvataju: multivarijantnu meta-analizu i meta-analitičko modelovanje strukturnim jednačinama, metodološka meta-analiza (Schulze,

2007), mrežna meta-analiza (Schwarzer i sar., 2015), meta-analiza individualnih podataka pacijenata (Stewart i Clarke, 1995), meta-analiza podataka iz ponovljenih merenja (Tanner-Smith i Tipton, 2014).

Specijalne meta-analitičke tehnike koje su namenjene sintezi psihometrijskih podataka su *generalizacija pouzdanosti* i *generalizacija valjanosti*. Pod pojmom *generalizacija pouzdanosti* podrazumeva se meta-analiza koeficijenata pouzdanosti. Ovaj izraz se sreće u studiji Kennedya i Turnagea (1991) u kojoj je urađena generalizacija test-retest koeficijenata pouzdanosti za devet kompjuterizovanih testova performansi. U studiji Vacha-Haase (1998) ponovo se uvodi pojam *generalizacije pouzdanosti*, sa namerom da se to prikaže kao nov meta-analitički postupak, koji nastaje kao proširenje metoda generalizacije valjanosti na generalizaciju koeficijenata pouzdanosti. Pod pojmom *generalizacija valjanosti* uglavnom se podrazumeva meta-analitički postupak razvijen od strane Hunter i Schmidt (Hunter i sar., 1982; Schmidt i Hunter, 2014). Ovim metodama se integrišu psihometrijski i statistički principi.

1.6.4. Modeli meta-analize

Dva glavna modela koji se koriste za kombinovanje rezultata studija su model fiksnih efekata i model slučajnih efekata.

Model fiksnog efekta - Ako je potvrđeno da između studija uključenih u meta-analizu ne postoji značajna heterogenost, onda možemo da izračunamo ukupnu ocenu svih podataka. Ovo je poznato kao model fiksnog efekta meta-analize. Objedinjena mera korišćena u fiksnom efektu meta-analiza će zavisi od ishoda interesa, kao što je RR, OR ili razlika srednjih vrednosti.

Model slučajnog efekta - Ako između studija postoji umerena ili značajna heterogenost potreban je alternativni pristup da se dobije ukupna ocena efekta a to je model slučajnog efekta meta-analize. Ovaj pristup podrazumeva u osnovi različitu snagu za svaku studiju koja je uključena u meta-analizu a to se uzima u obzir kao dodatni izvor varijacije.

Kod oba modela za procenu se koristi metod inverzne varijanse. Procenjuje se da je veličina efekta i njegova varijansa iz svake studije dovoljna da se prihvati metod inverzne varijanse (Schwarzer i sar., 2015).

Prilagodavanje modela podacima - Za model slučajnih efekata postoje različite metode za ocenu varijanse između studija: Hunter i Schmidt (Schmidt i Hunter, 2014), Hedges (Hedges i Olkin, 1993), DerSimonian-Laird (DerSimonian i Laird, 1986), Sidik-Jonkman (Sidik i Jonkman, 2005), maksimalne-verodostojnosti ili ograničene maksimalne-verodostojnosti (Viechtbauer, 2005), ili po Bayesu (Morris, 2012). U studijama sa ponovljenim merenjima za procenu varijanse koristi se metod robusne ocene varijanse³ (RVE) (Hedges i sar., 2010).

1.6.5. Procena publikacione pristrasnosti

Publikaciona pristrasnost je tendencija da se objavljuju studije na osnovu rezultata studije, a ne na osnovu njenog teorijskog ili metodološkog kvaliteta. Studije sa pozitivnim rezultatima imaju veću šansu da budu objavljene i citirane. Za procenu publikacione pristrasnosti koriste se grafičke i statističke metode (Rothstein i sar., 2006).

Kao grafički metod za identifikaciju postojanja publikacione pristrasnosti u izboru studija za sistematski pregled i meta-analizu koristi se levkasti dijagram⁴ (Sterne i Egger, 2001) samostalno ili uz ukloniti i popuniti metodu⁵ za procenu broja nedostajućih studija iz meta-analize zbog suzbijanja najekstremnijih rezultata na jednoj strani levkastog dijagrama (Duval i Tweedie, 2000a, 2000b).

Od statističkih metoda koje se koriste za detekciju asimetrije najčešće se primenjuje metod linearne regresije koji su predložili Egger i saradnici (1997). Ovaj metod procenjuje vezu između ocene efekta studije i njene standardne greške i može se smatrati statističkim analogom levkastog dijagrama. Dokaz o mogućoj publikacionoj pristrasnosti je $p < 0,05$.

³ *Robust Variance Estimation* (engl.)

⁴ *Funnel plot* (engl.)

⁵ *Trim and fill* (engl.)

1.6.6. Heterogenost

Studije koje se uključuju u meta-analizu treba da imaju sličan dizajn i da budu izabrane po zadatim kriterijumima. Međutim, neizbežno je da su studije u izvesnom stepenu heterogene. Heterogenost označava variranje opsega rezultata između studija uključenih u meta-analizu. Procene iz pojedinačnih studija mogu imati različite veličine ili čak različite pravce, pri čemu neke mogu da pokazuju povećani rizik a druge smanjeni rizik (Borenstein i sar., 2011; Schmidt i Hunter, 2014).

Varijabilitet ispitanika, intervencija i ispitivanih ishoda može se nazvati klinička heterogenost, razlika u dizajnu studija i rizik od publikacione pristrasnosti može se nazvati metodološka heterogenost (Higgins i Green, 2008). Klinička heterogenost podrazumeva: pacijent ili grupa bolesti (npr. pol, starost, vrsta zahvata), vreme ili intenzitet intervencije (npr. različite doze istog leka), zajednička intervencija (koji drugi tretman pacijent prima), merenje ishoda i vreme (Glasziou i Sanders, 2002). Heterogenost može da bude i u vezi sa nepoznatim ili nezabeleženim karakteristikama ispitivanja ili posledica slučajnosti. Prilikom objedinjavanja rezultata u meta-analizi važno je proceniti u kojoj meri se stvarni rezultati pojedinačnih studija razlikuju.

1.6.7. Analiza moderatora

Analiza moderatora izvodi se da se utvrdi izvor heterogenosti i koliko to doprinosi posmatranom varijabilitetu veličine efekta između studija. Bilo koja varijabla uključena u meta-analizu može uticati na veličinu efekta i može se smatrati potencijalnim moderatorom.

U literaturi je predloženo više metoda za sprovođenje analize moderatora (Glass, 1977; Hedges i Olkin, 1993; Rosenthal, 1991; Schmidt i Hunter, 2014). Postoje dva tipa regresionih modela, jedan je proširenje modela fiksnih efekata nazvan meta-regresioni model, a drugi je proširenje modela slučajnih efekata i zove se mešoviti model pošto obuhvata i fiksne i slučajne efekte. Meta-regresija modela fiksnih efekata je specijalni slučaj modela mešanih efekata kada je varijansa između studija $\tau^2=0$ (Schwarzer i sar., 2015). U savremenoj medicinskoj literaturi oba modela se zovu meta-regresija (Cooper i sar., 2009).

1.6.8. Analiza senzitivnosti

Analiza senzitivnosti zasnovana je na aspektima metodološkog kvaliteta i posebno je važna za meta-analizu. To je poslednja faza u sprovođenju analize i teži da ispita kolika je osetljivost rezultata meta-analize na uključene studije: različitih veličina, kvaliteta, karakteristika istraživanja ili faktora rizika koji se proučavaju, kao i drugih metodoloških razlika.

Analiza senzitivnosti može uključiti ponovnu analizu podgrupa originalnih podataka, kao i određivanje koliko pojedinačna studija (ili grupa studija sa sličnim karakteristikama) može uticati na ukupnu ocenu efekta. Jedna veoma velika studija može imati značajan, a možda i pogrešan (ako studija ima metodološki nedostatak), uticaj na ukupan rezultat. Analiza senzitivnosti može da pruži uvid u individualne faktore unutar studija koji mogu da utiču na rezultate. Procena uticaja metodološkog kvaliteta na ukupnu ocenu treba smatrati sastavnim delom procesa.

1.7. Meta-analiza ocena pouzdanosti i valjanosti

Podaci o pouzdanosti i valjanosti iz jedne studije ne mogu se automatski generalizovati na druge situacije jer na pouzdanost i valjanost mogu uticati različiti faktori: priroda uzorka, utreniranost i iskustvo ocenjivača, kao i uslovi u kojima se instrument primenjuje. Za dobijanje objedinjene ocene o psihometrijskim osobinama instrumenata neophodna je primena meta-analitičkih metoda. Broj meta-analiza ocena pouzdanosti skorova specifičnih testova raste rapidno u poslednjoj dekadi, najčešće pod nazivom generalizacija pouzdanosti (Botella i sar., 2010).

Saopštavanje objedinjenih ocena psihometrijskih osobina instrumenata ima pre svega praktičan značaj. Prilikom primene nekog testa u istraživanju najčešće se ne analiziraju njegove psihometrijske osobine već se citiraju iz prethodnih istraživanja. U toj situaciji istraživači mogu citirati samo onu literaturu gde su saopštene visoke ocene primenjenog instrumenta. Meta-analizom se izbegava takva pristrasnost, jer se u postupku dobijanja objedinjene ocene koriste svi podaci iz raspoložive literature.

U meta-analizi Stockings i sar. (2015) analizirane su skale za skrining simptoma za detekciju velikog depresivnog poremećaja kod dece i adolescenata. Za četiri primenjene skale (Reynoldova skala depresivnosti za adolescente, Bekova skala za procenu depresivnosti, Centar za epidemiološke studije – skala depresivnosti za decu i Dečiji upitnik za depresiju) interna pouzdanost je dobra, objedinjena vrednost 0,89 (95% CI 0,86 – 0,92). Objedinjena senzitivnost pomenutih skala iznosi 0,80 (95% CI 0,76 – 0,84) a specifičnost 0,78 (95% CI 0,74 – 0,83).

U meta-analizi generalizacije pouzdanosti Reynoldove skale depresivnosti za adolescente Koeficijent alfa imao je srednju vrednost od 0,90 sa opsegom od 0,70 do 0,97 (Vassar i Bradley, 2012).

U meta-analizama dijagnostičke adekvatnosti: Upitnik za poremećaj raspoloženja za skrining bipolarnih poremećaja, kod pacijenta sa poremećajima raspoloženja imao je objedinjenu senzitivnost 0,62 (95% CI 0,51 – 0,72) i specifičnost 0,85 (95% CI 0,79 – 0,89) (Wang i sar., 2015); Gerijatrijska skala depresije sa 30 ajtema je pokazala objedinjenu senzitivnost od 81,9% (95% CI 76,4 – 86,9) i specifičnost 77,7 (95% CI 73,0 – 82,1) (Mitchell i sar., 2010); U meta-analizi Tsai i sar. (2013) za Edinburg skalu postnatalne depresije, medijana koeficijenta alfa iznosila je 0,84. Za prag od ≥ 9 na skali, objedinjena vrednost senzitivnosti iznosi 0,94 (95% CI 0,68 – 0,99), a specifičnosti 0,77 (95% CI 0,59 – 0,88); Kod ispitanika sa HIV infekcijom, Centar za epidemiološke studije – skala za depresiju u detekciji velikog depresivnog poremećaja imala je objedinjenu senzitivnost 0,82 (95% CI 0,73 – 0,87) i specifičnost 0,73 (95% CI 0,63 – 0,80) (Tsai, 2014).

Shafer (2006) je prikazao meta-analizu faktorske strukture četiri instrumenta za procenu depresije, pri čemu je najčešći broj faktora za: Bekovu skalu za procenu depresivnosti iznosio 3 faktora (opseg, 2 do 7), Centar za epidemiološke studije – skalu za depresiju 4 faktora (opseg, 2 do 5), Hamiltonovu skalu za procenu depresivnosti 6 faktora (opseg, 3 do 8) i Zungovu skalu za samoprocenu depresije 3 faktora (opseg, 2 do 5). Srednje varijanse objašnjene faktorima iznosile su: 46%, 51%, 58% i 43%, respektivno.

1.8. Psihometrijske osobine Hamiltonove skale za procenu depresivnosti

Hamiltonova skala za procenu depresivnosti (HAMD) je jedna od prvih skala koja je razvijena da kvantifikuje težinu simptoma depresije kod ispitanika kojima je dijagnoza depresije već postavljena (Hamilton, 1967, 1960). Uvođenje ove skale dovelo je do promene u načinu istraživanja depresije. HAMD skala spada među najčešće korišćene skale u psihijatriji, uključujući primarnu zdravstvenu zaštitu, klinička istraživanja i kontrolisane kliničke ogleda (Bagby i sar., 2004). Postoji više verzija ove skale koje se razlikuju prema broju ajtema, interpretaciji ajtema i pravilima skorovanja (Williams, 2001), od kojih se u kliničkim istraživanjima verzija sa 17 ajtema koristi kao primarna skala (Licht i sar., 2005).

Originalna HAMD skala sa 17 ajtema primenjuje se kao strukturisani intervju u kojem se od ispitanika traži da odgovore na pitanja o znacima i simptomima tokom poslednje dve nedelje. Odgovori se ocenjuju od strane osobe koja vodi intervju. Skala sadrži 8 afektivno/kognitivnih ajtema (koji se ocenjuju sa 0 do 4) i 9 somatskih ajtema (koji se ocenjuju 0 do 3). Ukupni skor je u opsegu 0-52, pri čemu skor od 24 i više ukazuje na tešku depresiju (Doering i sar., 2007; Hamilton, 1960).

Među depresivnim skalama, HAMD skala se najčešće smatra “zlatnim standardom“ za procenu depresije (Carneiro i sar., 2015). HAMD skala procenjuje širok dijapazon simptoma depresije, što je i potvrđeno kroz multidimenzionalnost skale (Bech i sar., 2011).

U meta-analizi ocena pouzdanosti HAMD skale (Trajković i sar., 2011) dobijene su sledeće objedinjene vrednosti za: *internu konzistentnost*, koeficijent alfa 0,789 (95%CI 0,766-0,810), *međuposmatračku pouzdanost* (a) intraklasni koeficijent korelacije 0,937 (95%CI 0,914-0,954), (b) kappa koeficijent 0,81 (95%CI 0,72-0,88), (c) Pearsonov koeficijent korelacije 0,94 (95%CI 0,90-0,97), (d) Spearmanova korelacija rangova 0,91 (95%CI 0,78-0,96), i za *test-retest pouzdanost* između 0,65 i 0,98. Rezultati ukazuju da HAMD skala pruža pouzdanu procenu depresije.

U sistematskim pregledima valjanosti HAMD skale za period do 1979 (Hedlund i Vieweg, 1979) i period 1979-2004 (Bagby i sar., 2004) pokazano je da HAMD skala ima adekvatnu valjanost, kao i multidimenzionalnost.

U studiji Meader i sar. (2014) za skrining depresije kod pacijenata posle moždanog udara, HAMD skala je pokazala senzitivnost 0,84 (95% CI 0,75 - 0,90) i specifičnost od 0,83 (95% CI 0,72 - 0,90).

1.9. Psihometrijske osobine Montgomeri-Ašbergove skale za procenu depresivnosti

Montgomeri-Ašbergova skala za procenu depresivnosti (Montgomery i Asberg, 1979) konstruisana je i dizajnirana da bude senzitivnija na promene tokom tretmana u odnosu na HAMD skalu (Iannuzzo i sar., 2006). MADRS skala izvedena je iz pitanja koja obuhvataju depresivni sindrom iz jedinstvene psihijatrijske skale (CPRS⁶) (Asberg i sar., 1978).

Skalu primenjuju i popunjavaju psihijatri na osnovu razgovora - intervjuja vođenog sa pacijentom koji treba da omogući kompletno sagledavanje stanja kako bi se dobili precizni odgovori na postavljena pitanja. Skala sadrži 10 ajtema sa sedam stepeni od kojih parni (0, 2, 4, 6) imaju definisane odgovore, dok neparni (1, 3, 5) ocenu dobijaju interpolacijom parnih odgovora (Timotijević i Paunović, 2003).

U sistematskom pregledu instrumenata za merenje depresivnih simptoma kod pacijenata sa shizofrenijom Lako i sar. (2012) su za MADRS skalu prikazali objedinjene vrednosti interne konzistencije (0,91), međuposmatračke pouzdanosti (0,81) i test-retest pouzdanosti (0,71). Objedinjena srednja vrednost konkurentne valjanosti iznosila je 0,75 (0,56-0,90).

⁶ *Comprehensive Psychiatric Rating Scale (engl.)*

U studiji Meader i sar. (2014) za skrining depresije kod pacijenata posle moždanog udara, MADRS skala je pokazala senzitivnost 0,85 (95% CI 0,78 - 0,90) i specifičnost od 0,79 (95% CI 0,70 - 0,86).

Higuchi i sar. (2008) na osnovu objavljenih radova saopštavaju da se broj faktora kod MADRS skale kreće od 2 do 4, dok Uher i sar. (2008) smatraju da je MADRS unidimenzionalna skala.

1.10. Psihometrijske osobine Bekove skale za procenu depresivnosti

Bekova skala za procenu depresivnosti (BDI) (Beck i sar., 1961) je skala samoprocene, dizajnirana za ocenu stepena depresivnih simptoma koje ispitanik doživljava u tom trenutku (Cuijpers i sar., 2011). BDI je najšire korišćena skala samoprocene, koja ocenjuje kognitivne i afektivne karakteristike depresije (Brown i sar., 1995). Originalna verzija skale ima 21 pitanje sa četiri mogućnosti za odgovor stepenovan od 0 do 4. Ukupan skor na skali dobija se jednostavnim sabiranjem svih odgovora dobijenih od 1. do 21. pitanja (Timotijević i Paunović, 2003). Postoji skraćena verzija od 13 ajtema (Beck i Beck, 1972), revizija BDI-IA (Beck i sar., 1979) i revizija BDI-II (Beck i sar., 1996) urađena da skala bude u skladu sa DSM-IV kriterijumima.

BDI upitnik predstavlja zlatni standard za skale samoprocene depresije (Spielmans i sar., 2007). Skale samoprocene, kao što je BDI, imaju neke prednosti u odnosu na skale procene: uzimaju manje vremena, ne zahtevaju obučeno osoblje i njihova primena i ocenjivanje su više standardizovane (Cuijpers i sar., 2008).

1.11. Longitudinalne studije

Prema cilju istraživanja studije se dele na: opservacione i interventne (eksperimentalne). Opservacione studije mogu biti *deskriptivne* (ekološke studije, prikaz slučaja i serije slučajeva) i *analitičke* (anamnestičke (studije slučaj kontrola), kohortne (retrospektivne i prospektivne) i studije preseka. Eksperimentalne studije mogu biti (a) klinički ogled i (b) preventivni ili terenski ogled i (c) ogled u zajednici (Gordis, 2013; Guyatt, 2014; Machin i Campbell, 2005; Rothman i sar., 2008; Wassertheil-Smoller i Smoller, 2015; Woodward, 2013).

U odnosu na vreme prikupljanja podataka, neke studije prikupljaju podatke u sadašnjosti (studije preseka), iz prethodnog perioda (studije slučaj kontrola, retrospektivne kohortne studije) ili će podaci biti prikupljeni u budućem vremenu (prospektivne kohortne studije, randomizovani klinički ogled).

Postoji više tipova studija sa ponovljenim merenjima: longitudinalne studije, ukršteni eksperiment, "split-plot" eksperimenti sa ponovljenim merenjima u vremenu, "split-plot" eksperimenti sa ponovljenim merenjima na više lokacija. Pod longitudinalnim studijama podrazumevaju se studije sa ponovljenim merenjima istih varijabli na istim ispitanicima u različitim tačkama u vremenu (Ramsey i Schafer, 2012).

2. Ciljevi istraživanja

Longitudinalne studije su česte u psihijatriji, pri čemu se razne skale koriste za procenu statusa pacijenata i efikasnosti novih terapijskih procedura, pa je od izuzetne važnosti proučavanje psihometrijskih osobina ovih instrumenata pri ponovljenim merenjima (Laenen et al., 2009). Korišćenje pouzdanih i validnih instrumenata za procenu depresije je važno za kvantifikovanje promena u stepenu depresije tokom vremena. Ispitivanje promena korelacija između skala u toku vremena pomaže da se utvrdi konkurentna validnost ovih instrumenta u različitim tačkama u vremenu.

Uzimajući u obzir ograničenja prethodnih preglednih članaka i studija i neizvesnost promena ocena pouzdanosti i valjanosti u tokom vremena cilj ove studije bio je:

1. Uraditi sistematski pregled literature i meta-analitičko objedinjavanje ocena pouzdanosti i valjanosti psihijatrijskih skala saopštenih u različitim vremenima tokom longitudinalnih studija.
2. Proceniti promene ocena pouzdanosti i valjanosti psihijatrijskih skala saopštenih u različitim vremenima tokom longitudinalnih studija.
3. Generisati moderatore / prediktore prethodno navedenih ocena i njihovih promena.

Ovi ciljevi podrazumevali su analizu korisnosti meta-analize za kvantitativnu sintezu pouzdanosti i valjanosti psihijatrijskih skala u longitudinalnim studijama, i prikaz dostupnih meta-analitičkih metode za objedinjavanje publikovanih ocena o pouzdanosti valjanosti psihijatrijskih skala, kao i onih dobijenih sekundarnom analizom na osnovu publikovanih individualnih podataka.

Navedeni ciljevi bili su realizovani u studiji slučaja meta-analize pouzdanosti i valjanosti u longitudinalnim studijama dve često korišćene skale u psihijatriji: Hamiltonove skale za procenu depresivnosti i Montgomeri-Ašbergove skale za procenu depresivnosti.

Ove dve skale su odabrane jer su najkorišćeniji psihijatrijski instrumenti koji se koriste od strane ocenjivača. Obe se koriste za procenu depresivnosti, što je omogućilo meta-analizu publikovanih korelacija između njih, i ujedno su skale koje koriste ocenjivači, najčešće psihijatri, što je omogućilo i analizu međuposmatračke pouzdanosti.

Veliki broj publikovanih radova dao je potencijal za nalaženje dovoljnog broja ocena pouzdanosti i valjanosti Hamiltonove skale za procenu depresivnosti i Montgomeri-Ašbergove skale za procenu depresivnosti u longitudinalnim studijama koje su obuhvaćene sistematskim pregledom i meta-analizom.

3. Materijal i metod istraživanja

3.1. Pretraživanje literature, selekcija studija i ekstrakcija podataka

Jedinice posmatranja su literaturne jedinice (publikovani radovi) dobijene sistematskim pregledom literature u cilju identifikacije publikovanih istraživanja u kojima su saopštene ocene pouzdanosti i/ili valjanosti skala HAMD i MADRS.

Pretraživanje literature sprovedeno je u dve faze.

U **I FAZI**, pretražene su dve bibliografske (indeksno-abstraktne) baze podataka: MEDLINE (izdavač: U.S. National Library of Medicine) i PsycINFO (izdavač: American Psychological Association) korišćenjem EBSCO pretraživača koji omogućava pretraživanje po egzaktnim frazama. Pretraživanje je obavljeno kroz sva polja (uključujući i sažetke) ove dve baze sa zahtevom da se izdvoje sve jedinice u kojima se spominje HAMD i/ili MADRS skala.

Prilikom pretraživanja pomenutih baza podataka, za ocene pouzdanosti, faktorsku i dijagnostičku valjanost, u zahtev za pretraživanje unesene su fraze koje se odnose na:

- HAMD skalu: (hamilton rating scale for depression), (hamilton scale for depression), (hamilton depression rating scale), (hamilton depression scale) i
- MADRS skalu: (montgomery asberg depression rating scale), (montgomery asberg depression scale) i (montgomery and asberg depression rating scale)

Prilikom pretraživanja pomenutih baza podataka, za ocenu konkurentne valjanosti, u zahtev za pretraživanje unesene su i fraze koje se odnose i na Bekovu skalu za procenu depresivnosti (BDI):

- BDI: (Beck Depression Inventory)

Korišćene su različite kombinacije ovih fraza da se pronađu studije sa korelacijama između odgovarajućih psihijatrijskih skala.

Primarno pretraživanje, bez jezičkih restrikcija, sprovedeno je u periodu od 1. januara 2014. do 1. juna 2014. godine. Domaća literatura pretražena je ručno i obuhvata sve one časopise u kojima se objavljuju radovi iz psihijatrije i srodnih oblasti.

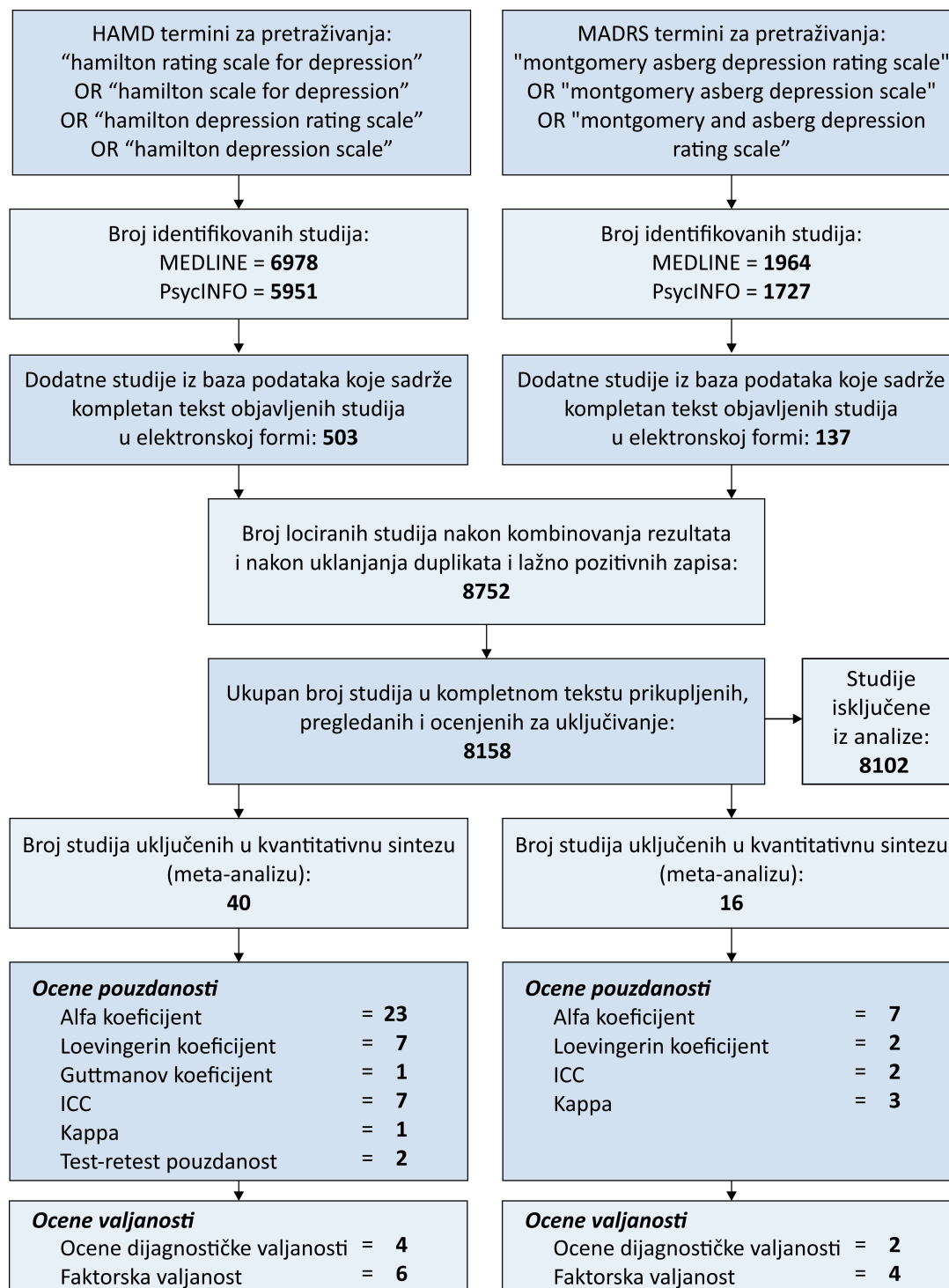
Vremenski period obuhvaćen pretraživanjem proteže se od početka primene skala HAMD (1960. godine) i MADRS (1979. godine) do kraja 2013 godine.

Uklonjeni su lažno pozitivni zapisi, u kojima nije korišćena niti HAMD skala niti MADRS skala. Rezultati pretraživanja obe baze podataka su objedinjeni kako bi se eliminisali duplikati. Po završenom pretraživanju bibliografskih baza podataka pristupilo se lociranju i prikupljanju radova u celini (In extenso).

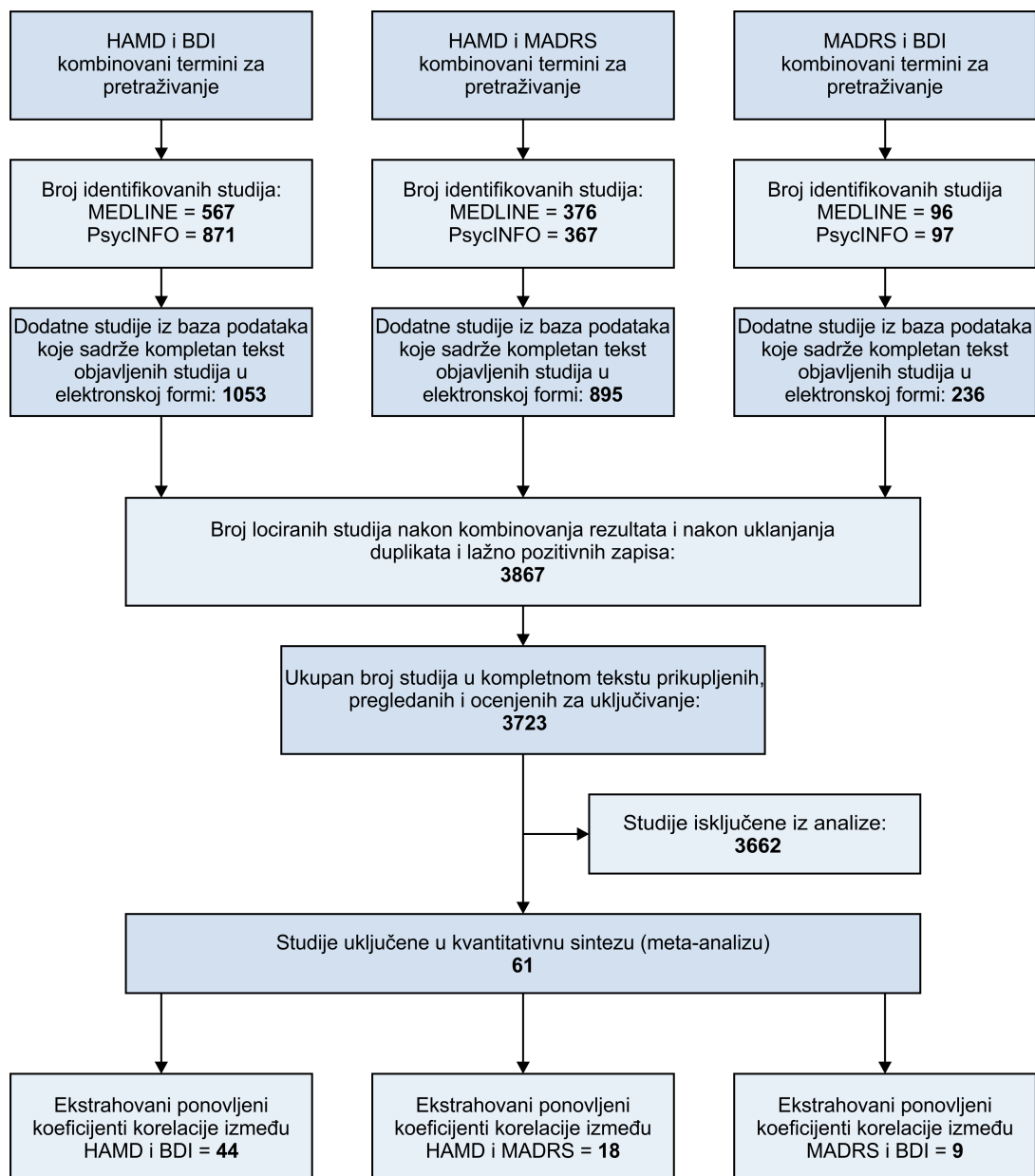
Ukupan broj prikupljenih radova u celini:

- za ocene pouzdanosti, dijagnostičke i faktorske valjanost iznosi 8158 (Slika 2)
- za ocene korelacija između skala HAMD, BDI i MADRS iznosi 3723 (Slika 3).

U **II fazi** izvršeno je pretraživanje ocena pouzdanosti i valjanosti u prikupljenim radovima u celini.



Slika 2. Dijagram toka strategije i rezultata pretraživanja za ocene pouzdanosti, faktorske i dijagnostičke valjanosti



Slika 3. Dijagram toka strategije i rezultata pretraživanja za koeficijente korelacija između skala HAMD, MADRS i BDI

3.2. Kriterijumi za uključivanje, neuključivanje i isključivanje studija iz meta-analize

Studije su smatrane podesnim za uključenje u dalju analizu ako zadovoljavaju sledeće uslove:

1. skale HAMD i MADRS su primenjene u jednoj od sledećih situacija:
 - u studijama gde su ove skale predmet istraživanja u cilju njihove evaluacije ili ispitivanja psihometrijskih karakteristika;
 - u studijama gde su ove skale korišćene za evaluaciju ili ispitivanja psihometrijskih karakteristika drugih psihijatrijskih instrumenata (skala, upitnika);
 - u studijama u kojima su ove skale korišćene u cilju ocene kliničkog stanja pacijenata;
2. saopštena je bar jedna ocena koja se odnosi na pouzdanost i valjanost skala HAMD ili MADRS i to u najmanje dva vremena.

Dodatni kriterijumi za uključivanje studija u analizu korelacija su sledeći: (1) saopšteni ponovljeni koeficijenti korelacije između skala HAMD, MADRS i BDI u longitudinalnom dizajnu ili saopšteni individualni podaci pacijenta¹ (IPD) na ovim skalama u longitudinalnom dizajnu, i (2) saopšteno vreme ocenjivanja depresije.

Kriterijum za uključenje u faktorsku analizu bio je da su saopštene faktorske strukture na bar dva merenja sa faktorskim opterećenjima.

Kriterijumi za uključenje studija u meta-analizu dijagnostičke valjanosti i njihovih promena u vremenu, da su mere Sn, Sp, LR⁺ i LR⁻ bile saopštene vezano za dijagnostičke performanse skala HAMD ili MADRS, a ne za predikciju u prognostičkim studijama.

U istraživanje nisu bile uključene studije u kojima HAMD i MADRS skale nisu primenjene na način kako su to predvideli autori ovih skala, kao ni studije koje nije bilo moguće locirati kao radove objavljene u celini.

Iz analize su isključene one studije koje ne ispunjavaju gore navedene kriterijume za uključivanje.

¹ *Individual patients data* (engl.)

3.3. Ekstrakcija podataka

Podaci su nezavisno ekstrahovani od strane dva istraživača (ZB i GT) u unapred pripremljen formular. Svi nezavisno ekstrahovani podaci su verifikovani za podudarnost, a sva neslaganja su rešena konsenzusom.

Ekstrakcija i kodiranje podataka obavljeno je na sledeći način:

1. Podaci o pouzdanosti i valjanosti za skale HAMD i MADRS u primarnim studijama

1.1. Koeficijenti pouzdanosti

- 1.1.1. Alfa koeficijent
- 1.1.2. Loevingerin koeficijent
- 1.1.3. Koeficijent pouzdanosti dve polovine
- 1.1.4. Koeficijent međuposmatračke pouzdanosti - ICC
- 1.1.5. Koeficijent međuposmatračke pouzdanosti - kappa
- 1.1.6. Koeficijent test-retest pouzdanosti ukupnih skorova

1.2. Koeficijenti valjanosti

- 1.2.1. Koeficijenti korelacija
- 1.2.2. Ocene dijagnostičkih performansi
 - 1.2.2.1. Senzitivnost i specifičnost
 - 1.2.2.2. Pozitivna i negativna prediktivna vrednost
 - 1.2.2.3. Površina ispod krive maksimalne verodostojnosti² (AUC)
- 1.2.3. Podaci iz faktorskih analiza
 - 1.2.3.1. Korelaciona matrica
 - 1.2.3.2. Broj ekstrahovanih faktora
 - 1.2.3.3. Ekstrahovana varijansa u faktorskoj analizi
 - 1.2.3.4. Faktorska matrica (matrica faktorskih opterećenja)

² *Area under curve* (engl.)

2. Karakteristike studije
 - 2.1. Godina publikovanja
 - 2.2. Tip studije
 - 2.2.1. Metodološka studija (psihometrijska studija gde je cilj ispitivanja HAMD ili MADRS ili neka druga psihijatrijska skala, studija o statističkom metodu ili studija o dijagnostičkoj valjanosti)
 - 2.2.2. Klinička studija
 - 2.3. Veličina uzorka
3. Karakteristike ispitanika
 - 3.1. Starost ispitanika
 - 3.2. Lokacija ispitanika
 - 3.2.1. Hospitalizovani pacijenti
 - 3.2.2. Ambulantni pacijenti
 - 3.2.3. Hospitalizovani + ambulantni pacijenti
 - 3.3. Jezičko područje
 - 3.3.1. Engleski jezik
 - 3.3.2. Ostali jezici
 - 3.4. Dijagnostička grupa ispitanika
 - 3.4.1. Depresija kao primarna dijagnoza
 - 3.4.2. Ostalo
 - 3.5. Skor na HAMD skali za svaku procenu
 - 3.6. SD skora na HAMD skali za svaku procenu
 - 3.7. Skor na MADRS skali za svaku procenu
 - 3.8. SD skora na MADRS skali za svaku procenu
 - 3.9. Skor na BDI skali za svaku samoprocenu
 - 3.10. SD skora na BDI skali za svaku samoprocenu
4. Karakteristike primenjene verzije HAMD skale
 - 4.1. Broj ajtema na HAMD skali
 - 4.2. Modifikacija HAMD skale
 - 4.2.1. Hamilton (1960, 1967)
 - 4.2.2. Ostale modifikacije

5. Karakteristike primenjene verzije MADRS skale
 - 5.1. Broj ajtema na MADRS skali
 - 5.2. Modifikacija MADRS skale
 - 5.2.1. Montgomery-Åsberg (1979)
 - 5.2.2. Ostale modifikacije

6. Karakteristike primenjene verzije BDI skale
 - 6.1. Broj ajtema na BDI skali
 - 6.2. Modifikacija BDI skale
 - 6.2.1. Beck (1961)
 - 6.2.2. Ostale modifikacije

7. Karakteristike mernog procesa i načina izračunavanja koeficijenata
 - 7.1. Veličina uzorka na kojem je dobijena ocena pouzdanosti i (ili) valjanosti
 - 7.2. Dijagnostički kriterijum
 - 7.2.1. Istraživački dijagnostički kriterijumi³ (RDC)
 - 7.2.2. Dijagnostičko i statističko uputstvo mentalnih poremećaja⁴ DSM-III
 - 7.2.3. DSM-III-R
 - 7.2.4. DSM-IV
 - 7.2.5. Međunarodna klasifikacija bolesti, deseta revizija⁵ ICD-10
 - 7.2.6. Ostalo
 - 7.3. Obrazovanje ocenjivača
 - 7.3.1. Psihijatri i klinički psiholozi
 - 7.3.2. Ostali
 - 7.4. Način izračunavanja međuposmatračke pouzdanosti
 - 7.4.1. ICC
 - 7.4.2. Pearsonov r
 - 7.4.3. Spearmanov ρ
 - 7.4.4. Kappa koeficijent
 - 7.4.5. Kendallov W

³ *Research Diagnostic Criteria* (engl.)

⁴ *Diagnostic and Statistical Manual of Mental Disorders* (engl.)

⁵ *International Classification of Diseases* (engl.)

- 7.5. Način izračunavanja koeficijenta korelacije
 - 7.5.1. Pearsonov r
 - 7.5.2. Spearmanov ρ
- 7.6. Način na koji je sprovedena procena ispitanika u toku intervjua
 - 7.6.1. Zajednički intervjui
 - 7.6.2. Individualni intervjui ocenjivača
 - 7.6.3. Video zapis
 - 7.6.4. Audio zapis
- 7.7. Vreme ocenjivanja (dani) za svaku procenu
- 7.8. Vreme između dva intervjua u slučaju test-retest pouzdanosti

3.4. Analiza podataka

Analiza podataka ocena pouzdanosti i valjanosti u ovom istraživanju izvršena je korišćenjem više meta-analitičkim metoda, usled postojanja više tipova pouzdanosti i valjanosti, i specifičnosti podataka dobijenih iz ponovljenih merenja.

Pre same analize svi koeficijenti pouzdanosti i koeficijenti korelacije su zed-transformisani. Vreme procene je logaritamski transformisano zbog pozitivno iskošenih podataka. Primenjene analize u odnosu na tip koeficijenta pouzdanosti i valjanosti prikazane su u Tabeli 1.

Objedinjavanje podataka rađeno je na dva načina: (a) objedinjavanjem zed transformisanih vrednosti koeficijenata svih studija na svim merenjima koristeći metod robusne ocene varijanse (Hedges i sar., 2010) i (b) objedinjavanjem zed transformisanih koeficijenata na prvoj i poslednjoj proceni koristeći metode Hedgesa i Olkina (1993), uz prilagođavanje modela podacima prema DerSimonian i Lairdu (1986).

Analiza moderatora izvedena je u meta-analizi: (1) Alfa koeficijenata za HAMD skalu, (2) Loevingerinih koeficijenata za HAMD skalu, (3) ICC za HAMD skalu, (4) korelacija između skala HAMD i BDI, kao i između skala HAMD i MADRS.

Analiza heterogenosti svih studija u svim merenjima izvedena je u okviru robumeta paketa izračunavanjem τ^2 i I^2 , dok je na prvom i poslednjem merenju za analizu heterogenosti korišćena Q statistika i I^2 .

Analiza senzitivnosti izvedena je sa različitim vrednostima ρ u okviru paketa robumeta i analizom podgrupa.

Jedinica analize bila je individualna studija.

Hipoteze su testirane na nivou statističke značajnosti (alfa nivo) 0,05.

Metodologija primenjena u ovom istraživanju je u skladu sa PRISMA (Moher i sar., 2009) i MOOSE (Stroup i sar., 2000) preporukama i pravilima.

Tabela 1. Prikaz analiza primenjenih u istraživanju

Meta-analički metod	Alfa koeficijent		Loevingerin koeficijent		ICC		Kappa	Test-retest	Korelacije			Dijagnostička valjanost	Faktorska valjanost
	HAMD	MADRS	HAMD	MADRS	HAMD	MADRS	MADRS	HAMD	HAMD	BDI	BDI	HAMD	HAMD
Objedinjena vrednost													
-prva procena	+	+	+	+	+	+	+	+	+	+	+	+	+
-poslednja procena	+	+	+	+	+	+	+	+	+	+	+	+	+
Objedinjena regresiona linija	+	+	+		+				+	+	+		
Forest plot													
-prva procena	+	+	+	+	+	+		+	+	+	+	+	+
-poslednja procena	+	+	+	+	+	+		+	+	+	+	+	+
Meta- regresija													
-univarijantna	+		+		+				+	+			
-multivarijantna	+								+				
Analiza senzitivnosti													
-različite vrednosti p	+		+						+	+	+		
-analiza podgrupa									+				
Levkasti dijagram													
-prva procena	+	+	+		+				+	+	+		
-poslednja procena	+	+	+		+				+	+	+		

HAMD: Hamiltonova skala za procenu depresivnosti; MADRS: Montgomeri-Ašbergova skala za procenu depresivnosti; BDI: Bekova skala za procenu depresivnosti; ICC: Intraklasni koeficijent korelacije

3.4.1. Korekcija i transformacija podataka

Korekcija podataka izvedena je za alfa koeficijente pre transformacije primenom Spearman-Brown formule. Transformacija podataka izvedena je Fisherovom z-transformacijom za korigovane alfa koeficijente, koeficijente međuposmatračke pouzdanosti, koeficijente test-retest pouzdanosti i koeficijente korelacije.

3.4.1.1. Spearman-Brownova korekcija

Poznato je da na pouzdanost instrumenata utiče i njegova dužina. Sa porastom broja ajtema raste i interna konzistencija (Fajgelj, 2013) pa su sve vrednosti alfa koeficijenta korigovane Spearman-Brownovom formulom:

$$r_c = \frac{kr_o}{1 + (k - 1)r_o}$$

gde je r_c korigovana vrednost alfa koeficijenta, k faktor skraćenja (odnos broja ajtema skale na čiji nivo treba korigovati vrednosti i aktuelnog broja ajtema), a r_o je nekorigovana vrednost alfa koeficijenta.

3.4.1.2. Fisherova z-transformacija

Pre uključivanja u analizu koeficijenti pouzdanosti i korelacije su podvrgnute Fisherovoj (1921) z-transformaciji:

$$z = \frac{1}{2} \ln \frac{1+r}{1-r},$$

gde je r koeficijent pouzdanosti ili koeficijent korelacije.

Po završenoj analizi transformisanih vrednosti, zbog lakše interpretacije, rezultati su obrnutom transformacijom

$$r = \frac{e^{2z_r} - 1}{e^{2z_r} + 1}$$

dovedeni na nivo originalnih koeficijenata pouzdanosti.

3.4.2. Robusna ocena varijanse u meta-regresiji sa veličinom efekta kao zavisnom varijablom

Glavni problem u meta-analizi podataka sa ponovljenih merenja istog konstrukta koje se odnose na iste individue je nesaopštavanje strukture kovarijanse u publikovanim studijama. Statističkom tehnikom robusne ocene varijanse (Hedges i sar., 2010) prevazilazi se nedostatak matrice kovarijanse. Ovom tehnikom izračunava se standardna greška na osnovu empirijske ocene varijanse. Ona ne zahteva bilo kakvu pretpostavku o raspodeli ocena veličine efekta koje su predmet meta-analize. U simulacionim studijama pokazano je da je metod prihvatljiv sa dostupnošću od najmanje 10 studija za ocenu prosečne veličine efekta ili 20-40 studija za ocenu nagiba. Važna korist koja se dobija primenom ovog metoda jeste i to da su uključuje maksimalan broj studija u meta-analizu jer metod zahteva minimalan broj publikovanih podataka. Prednost ovog metoda u odnosu na analizu razlika na prvoj i poslednjoj proceni koeficijenata pouzdanosti ili valjanosti, uz korišćenje Monte Carlo simulacija za procenu standardne greške tih promena, je u tome da ona omogućava analizu više prediktora iz svih merenja.

Metod pretpostavlja model slučajnih efekata u kojima prosečne veličine efekta variraju između različitih studija. Metod RVE može se koristiti za bilo koji tip podataka sa statističkom zavisnošću. Uopšteno postoje dva tipa statističke zavisnosti koji se mogu javiti u meta-analizi: korelisani efekti i hijerarhijski efekti (Tipton, 2014). Korelisani efekti javljaju se u slučaju kada studija saopštava ponovljene veličine efekta koji se odnose na iste jedinice posmatranja-dizajn ponovljenih merenja. U ovakvim studijama retko su publikovane ocene korelacija zavisnosti podataka, što otežava ili čini nemogućim primenu multivarijantnog meta-analitičkog pristupa. Drugi tip statistički zavisnih podataka u meta-analizi jesu hijerarhijski podaci, što podrazumeva višestruko saopštene veličine efekta koje se ne odnose na iste jedinice posmatranja u studiji, već na različite uzorke ispitanika unutar iste studije, uključujući i multicentrične studije.

U metodu robusne ocene varijanse odnos vektora ocene veličine efekta (\mathbf{T}), matrice vrednosti prediktorskih varijabli (\mathbf{X}), vektora regresionih koeficijenata ($\boldsymbol{\beta}$) i vektora grešaka ($\boldsymbol{\varepsilon}$) je definisan formulom

$$\mathbf{T} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Regresioni koeficijenti ocenjuju se metodom ponderisanih najmanjih kvadrata.

$$\mathbf{b} = \left(\sum_{j=1}^m \mathbf{X}'_j \mathbf{W}_j \mathbf{X}_j \right)^{-1} \left(\sum_{j=1}^m \mathbf{X}'_j \mathbf{W}_j \mathbf{T}_j \right)$$

za pojedinačne studije $j=1 \dots m$, a \mathbf{W} je matrica pondera.

Ocena uzoračke varijanse regresionih koeficijenata metodom RVE dobija se po formuli:

$$\mathbf{v}^R = \left(\sum_{j=1}^m \mathbf{X}'_j \mathbf{W}_j \mathbf{X}_j \right)^{-1} \left(\sum_{j=1}^m \mathbf{X}'_j \mathbf{W}_j \mathbf{e}_j \mathbf{e}'_j \mathbf{W}_j \mathbf{X}_j \right) \left(\sum_{j=1}^m \mathbf{X}'_j \mathbf{W}_j \mathbf{X}_j \right)^{-1}$$

gde je $\mathbf{e}_j = \mathbf{T}_j - \mathbf{X}_j \mathbf{b}$ vektor reziduala u j -toj studiji.

Hipoteza da su regresioni koeficijenti jednaki nuli ($\beta_k = 0$) može biti testirana Waldovom statistikom:

$$t_k^{Rs} = \frac{b_k}{\sqrt{V_{kk}^{Rs}}} = \frac{b_k}{\sqrt{\left(\frac{m}{m-p} \right) V_{kk}^R}}$$

gde je m broj studija, p je broj parametara koji se ispituje, a $m-p$ je broj stepena slobode.

U modelu korelisanih efekata, težinski koeficijent se može izračunati po formuli:

$$w_{ij} = \frac{1}{\{k_j (V_{\bullet j} + \tau^2)\}}$$

prihvatanjem da je za svaku studiju j : korelacija veličine efekta između svih parova konstanta ρ , ove korelacije su iste u svim studijama i k_j uzoračka varijansa unutar studija je približno ista, sa prosečnom varijansom $V_{\bullet j}$.

U modelu korelisanih efekata τ^2 se dobija preko formule:

$$\tau^2 = \frac{Q_E - m + \text{tr} \left[\mathbf{V} \left(\sum_{j=1}^m \frac{w_j}{k_j} \mathbf{X}'_j \mathbf{X}_j \right) \right] + \rho \text{tr} \left[\mathbf{V} \left(\sum_{j=1}^m \frac{w_j}{k_j} [\mathbf{X}'_j \mathbf{J}_j \mathbf{X}_j - \mathbf{X}'_j \mathbf{X}_j] \right) \right]}{\sum_{j=1}^m k_j w_j - \text{tr} \left[\mathbf{V} \left(\sum_{j=1}^m w_j^2 \mathbf{X}'_j \mathbf{J}_j \mathbf{X}_j \right) \right]}$$

u formuli \mathbf{J}_j predstavlja $k_j \times k_j$ matricu, $\mathbf{V} = (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1}$

Pokazano je da metod RVE ima dobre performanse kada je broj studija uključenih u meta-analizu veliki. Kada je broj studija mali Tipton (2014) je predložila korekciju za male uzorke koja uključuje rezidualne i broj stepena slobode u oceni RVE.

Opciono, bilo je moguće analizirati razliku na prvoj i poslednjoj proceni koeficijenta pouzdanosti ili valjanosti uz korišćenje Monte Carlo simulacija za procenu standardne greške tih promena. Aktuelno primenjeni metod meta-regresije sa RVE je u analitičkom smislu napredniji jer omogućava analizu više prediktora iz svih merenja.

3.4.3. Meta-analitički metod Hedgesa i Olkina

Meta-analitički metod Hedgesa i Olkina (1993) je metod približnog objedinjavanja podataka sa testiranjem homogenosti. Za testiranje homogenosti veličine efekata između studija uzima u obzir uzoračku varijansu svake individualne veličine efekta. Meta-analiza je izvedena u *modelu sa fiksnim efektima* i *modelu sa slučajnim efektima*.

3.4.3.1. Model sa fiksnim efektima

U modelu sa fiksnim efektima objedinjena aritmetička sredina transformisanih vrednosti koeficijenta pouzdanosti dobijena je korišćenjem težinskih koeficijenta:

$$\bar{T} = \frac{\sum w_i T_i}{\sum w_i}$$

gde je T transformisana vrednost koeficijenta pouzdanosti, a w_i je težinski koeficijent koji se dobija kao recipročna vrednost uzoračke varijanse:

$$w_i = \frac{1}{v_i}$$

gde je v_i uzoračka varijansa pojedinačne studije.

3.4.3.2. Model sa slučajnim efektima

U modelu sa slučajnim efektima analiza je izvedena na sledeći način:

- a) Izvedena je analiza po modelu fiksnih efekata da bi se odredile objedinjene statistike.
- b) Za procenu varijanse između studija kod modela slučajnih efekata najčešće se koristi metod DerSimonian i Lairda (1986), što je ujedno i najcitiraniji rad iz oblasti meta-analize (Shadish i Lecy, 2015). Model je izveden za izjednačavanje očekivane vrednosti Cochran Q statistike sa njegovom opserviranom vrednošću. Najčešća DerSimonian i Laird ocena τ^2 je prema formuli:

$$\tau^2 = \frac{Q - (k - 1)}{\sum w_i - \left(\frac{\sum w_i^2}{\sum w_i} \right)}$$

gde je Q statistika heterogenosti, pri čemu je $\tau^2 = 0$ ako je $Q < k - 1$

- c) Težinski koeficijenti w_i za model sa slučajnim efektima izračunavaju se:

$$w_i = \frac{1}{v_i + \tau^2}$$

- d) Tako dobijeni težinski koeficijenti koriste se u procesu izračunavanja koji je identičan onome u modelu sa fiksnim efektima.

3.4.4. Analiza podataka o dijagnostičkoj valjanosti

3.4.4.1. Rekonstrukcija dijagnostičke tabele

Za uključivanje studije u analizu dijagnostičke valjanosti neophodni su podaci na osnovu kojih je moguća rekonstrukcija dijagnostičke 2x2 tabele. Na osnovu podataka o ukupnom broju ispitanika, broja ispitanika kojima je postavljena dijagnoza depresije na osnovu kriterijumskog standarda, senzitivnosti i specifičnosti rekonstruisane su učestalosti iz dijagnostičke tabele: *tačno pozitivni (TP)*, *lažno negativni (LN)*, *lažno pozitivni (LP)* i *tačno negativni (TN)*.

Ukoliko nisu saopštene u studiji, iz rekonstruisane dijagnostičke tabele izračunate su druge ocene dijagnostičkih performansi: *Pozitivna prediktivna vrednost (PPV)*, *Negativna prediktivna vrednost (NPV)*, *Odnos verodostojnosti pozitivnog testa (LR+)*, *Odnos verodostojnosti negativnog testa (LR-)* i *Dijagnostički odnos šansi (DOR)*.

3.4.4.2. Objedinjavanje ocena dijagnostičkih performansi

Objedinjene vrednosti senzitivnosti i specifičnosti dobijene su po formulama:

$$Se_T = \frac{\sum TP}{\sum (TP + LN)} \quad Sp_T = \frac{\sum TN}{\sum (LP + TN)}$$

gde su TP , LN , TN i LP učestalosti iz dijagnostičke tabele. Formule odgovaraju ponderisanim aritmetičkim sredinama sa veličinom uzoraka kao težinskim koeficijentima.

Odnosi verodostojnosti i dijagnostički unakrsni odnos objedinjeni su metodom DerSimonian i Lairda (1986), odnosno po modelu slučajnih efekata, na osnovu sledeće formule:

$$\ln \theta_r^{DL} = \frac{\sum w_i^{DL} \ln \theta_i}{\sum w_i^{DL}}$$

gde je θ veličina efekta u studiji (LR+, LR- ili DOR), w_i^{DL} težinski koeficijent:

$$w_i^{DL} = \frac{1}{SE(\ln \theta_i)^2 + \tau^2}$$

3.4.4.3. Testiranje homogenosti ocena dijagnostičkih performansi

Homogenost senzitivnosti i specifičnosti je testirana primenom testa odnosa verodostojnosti (Agresti, 2010), koji sledi hi-kvadrat raspodelu sa $k-1$ stepena slobode (gde je k broj studija).

Efekat dijagnostičkog praga testiran je izračunavanjem Spearmanovog koeficijenta korelacije senzitivnosti i specifičnosti. Uzeto je da korelacija manja od $-0,60$ (tj. u intervalu od $-1,00$ do $-0,60$) ukazuje na postojanje dijagnostičkog praga (Devillé i sar., 2002; Moses i sar., 1993).

3.4.5. Procena publikacione pristrasnosti

Publikaciona pristrasnost u izboru studija za meta analizu procenjivana je na osnovu izgleda levkastog dijagrama sa iseci i popuni metodom za dodavanje nedostajućih studija (Duval i Tweedie, 2000a, 2000b; Sterne i Egger, 2001). Testiranje publikacione pristrasnosti rađeno je Eggerovim testom (Egger i sar., 1997).

3.4.6. Heterogenost

Heterogenost studija se može proceniti grafički preko forest plotu i Baujat dijagrama, kao i testiranjem heterogenosti (Q test i I^2).

Na **forest plotu** moguće je proceniti heterogenost. Izvuče se normala iz višeg kraja 95% intervala poverenja za studiju sa najmanjom veličinom efekta. Sličnu liniju povući kroz donji kraj 95% intervala poverenja studije sa najvišom veličinom efekta. Ako se intervali poverenja svih studija pojavljuju u prostor između ove dve linije, studije verovatno nisu heterogene. Svaka studija izvan ove oblasti može biti uzrok značajne heterogenosti u ukupnoj analizi rezultata studije.

Baujatov dijagram omogućava identifikaciju studija koje su izvor heterogenosti i kvantifikovanje doprinosa ovih studija na ukupan rezultat. Svaka studija je predstavljena tačkom ili rednim brojem na dijagramu. Na x-osi predstavljen je doprinos studije ukupnom Cochran Q testu heterogenosti. Na y-osi nalazi se uticaj studije, definisan kao standardizovani kvadrat razlike između efekta tretmana sa ili bez studije. Na x-osi se može naneti i kvadriran Pearsonov rezidual, a y-osa odgovara standardizovanom kvadratu razlike između predviđene vrednosti studije sa i bez uključanja studije u model. Studije koje se nalaze u gornjem desnom kvadrantu dijagrama najviše doprinose heterogenosti i ukupnim rezultatima (Baujat i sar., 2002).

Test heterogenosti za procenu postojanja statistički značajne heterogenosti između studija uključenih u meta-analizu je verzija hi-kvadrat testa poznata kao **Cochran- Q test** (Cochran, 1954; Hedges i Olkin, 1993). Testom se ispituje da li se

pojedinačni efekti razlikuju od zbirnog efekta više nego što bi smo očekivali na bazi slučajnosti. Test se izračunava po formuli:

$$Q = \sum w_i (T_i - \bar{T})^2$$

Izračunavanjem I^2 statistike moguće je kvantifikovati količinu statističke heterogenosti između studija. To se radi po formuli:

$$I^2 = \frac{Q - (k - 1)}{Q} \times 100$$

gde je Q - statistika za test heterogenosti a k - broj studija uključenih u meta-analizu.

Vrednosti statistike I^2 se nalaze u opsegu između 0 i 100 %. Za interpretaciju rezultata, vrednost 0% pokazuje da ne postoji posmatrana heterogenost a vrednostima I^2 statistike od 25%, 50% i 75% odgovara niska, srednja i visoka heterogenost.

Varijansa između studija (τ^2), najčešće izračunava metodom po DerSimonian i Lairdu. Formula za izračunavanje, navedena je u tekstu iznad.

3.4.7. Analiza moderatora

Analiza moderatora izvedena je primenom univarijantne i multivarijantne meta-regresije. Analiza moderatora izvedena je RVE metodom u *robumeta* programskom dodatku za R. Prilagođavanje metoda podacima izvršeno je po metodi DerSimonian i Lairda. Sve varijable potencijalni moderatora uključeni su u univarijantnu i multivarijantnu mera-regresionu analizu.

Univarijantna meta-regresija izvedena je za ocene: (1) Alfa koeficijenta za HAMD skalu, (2) Loevingerinog koeficijenta za HAMD skalu, (3) ICC za HAMD skalu, (4) korelacija između skala HAMD i BDI, kao i između HAMD i MADRS. Zbog numeričkih ograničenja RVE metoda i nepovoljnog odnosa između broja prediktora i broja studija multivarijantna meta-regresija izvedena je samo za ocene (1) Alfa koeficijenta za HAMD skalu i (2) korelacije skala HAMD i BDI.

3.4.8. Analiza senzitivnosti

Analiza senzitivnosti izvedena je sa različitim vrednostima ρ ($0 \leq \rho \leq 1$) u RVE metodi (Tanner-Smith i Tipton, 2014) i analizom podgrupa.

3.4.9. Softver korišćen za analizu podataka

Analiza podataka rađena je u R programskom jeziku i okruženju (R Core Team, 2014). R je idealan softverski paket za izvođenje meta-analize jer je slobodno dostupan i skripte mogu lako da se dele i reprodukuju (Quintana, 2015).

Objedinjavanje vrednosti zed transformisanih koeficijenata svih studija u vremenu za Alfa koeficijent, Loevingerin koeficijent, ICC, korelacije između skala HAMD i BDI, HAMD i MADRS, kao i za MADRS i BDI rađeno je u *robumeta* R programskom dodatku (Fisher i Tipton, 2015).

Objedinjavanje zed transformisanih vrednosti na prvoj i poslednjoj proceni za Alfa koeficijent, Loevingerin koeficijent, ICC, Kappa koeficijent i Test-retest koeficijent rađeno je u *metafor* R programskom dodatku (Viechtbauer, 2010). Objedinjavanje zed transformisanih vrednosti na prvoj i poslednjoj proceni za analizu korelacija između skala HAMD, MADRS i BDI rađeno je u *metafor* i *meta* (Schwarzer, 2014) R programskim dodacima. Objedinjavanje vrednosti dijagnostičke valjanosti na prvoj i poslednjoj proceni rađeno je u *mada* R programskom dodatku (Doebler, 2015) i Meta-DiSc softveru (Zamora i sar., 2006).

Za grafičko prikazivanje podataka korišćeni su *ggplot2* (Wickham, 2009) i *lattice* (Sarkar, 2008) R programski dodaci.

RSkripte sa komandama za analize upotrebljene u ovom istraživanju date su u Prilogu 1.

4. Rezultati

Pretraživanjem bibliografske baze podataka MEDLINE, za period od nastanka skala HAMD i MADRS do kraja 2013. godine, identifikovano je 6978 studija u kojima je korišćena HAMD skala i 1964 studija u kojima je korišćena MADRS skala. Za isti period, pretraživanjem bibliografske baze podataka PsycINFO, identifikovano je 5951 studija u kojima je korišćena HAMD skala i 1727 studija u kojima je korišćena MADRS skala. Rezultati pretraživanja obe baze su spojeni, a zatim uklonjeni duplikati koji se javljaju pošto je jedan broj časopisa indeksiran u obe baze. Ukupan broj identifikovanih studija iz obe baze, za period od nastanka skala do kraja 2013. godine, iznosi 7154 za HAMD skalu i 2143 za MADRS skalu. Pretraživanjem baza podataka koje sadrže kompletan tekst objavljenih studija u elektronskoj formi (*HINARI, EBSCO, ProQuest i Science Direct*) identifikovano je i locirano dodatnih 503 studija za HAMD skalu i 137 za MADRS skalu. Nakon spajanja baza i uklanjanja duplikata i lažno pozitivnih rezultata ukupan broj identifikovanih i lociranih studija iznosi 7314 studija za HAMD skalu i 2187 za MADRS skalu.

Ukupan broj prikupljenih studija u punom tekstu u kojima je korišćena HAMD skala iznosio je 6590 (90,1% od ukupnog broja lociranih studija) i 2051 za MADRS skalu (93,8% od ukupnog broja lociranih studija).

Studije koje nisu locirane u potpunom tekstu najčešće su iz časopisa sa niskim impakt faktorom ili bez impakt faktora, van engleskog govornog područja ili ne postoje u elektronskim bazama podataka i bibliotekama.

Locirane primarne studije bile su pregledane od dva nezavisna istraživača (ZB i GT) koji su ekstrahovali podatke, u unapred pripremljen formular, koji se odnose na ocene pouzdanosti i valjanosti HAMD skale i podatke koji se potencijalno mogu upotrebiti u analizi varijabli moderatora. Svi nezavisno ekstrahovani podaci su verifikovani za podudarnost, a sva neslaganja su rešena konsenzusom.

Od 6590 pregledanih studija za HAMD skalu 98 (1,5%) je identifikovano kao studije u kojima su saopšteni podaci o pouzdanosti ili valjanosti. Od 2051 pregledanih

studija za MADRS skalu 38 (1,9%) je identifikovano kao studije u kojima su saopšteni podaci o pouzdanosti ili valjanosti.

U Tabeli 2 prikazan je broj studija iz kojih su ekstrahovani podaci različitih tipova ocena pouzdanosti i valjanosti u longitudinalnim studijama za skale HAMD i MADRS. U najvećem broju studija saopšteni su podaci o korelacijama između skala (61 studija), a sledi broj studija u kojima su saopšteni alfa koeficijenti (29 studija).

Podatak o ukupnom broju studija uključenih u kvantitativnu sintezu, koji iznosi 111, ne može se dobiti prostim sabiranjem učestalosti za pojedine tipove ocena, jer je u nekim studijama saopšteno više od jednog tipa ocene pouzdanosti ili valjanosti.

Tabela 2. Broj studija iz kojih su ekstrahovani podaci

	HAMD	MADRS	Ukupno
<i>Ocene pouzdanosti</i>			
Alfa koeficijent	23	7	29
Loevingerin koeficijent	7	2	7
Guttmanov test pouzdanosti dve polovine	1	-	1
ICC	7	2	8
Kappa	1	3	3
Test-retest pouzdanost ukupnih skorova	2	-	2
<i>Ocene valjanosti</i>			
Korelacija HAMD i BDI	44	-	44
Korelacija HAMD i MADRS	18	18	18
Korelacija MADRS i BDI	-	9	9
Ocene dijagnostičke valjanosti	4	2	5
Faktorska valjanost	6	4	9
Ukupno studija iz kojih su ekstrahovane ocene pouzdanosti i valjanosti	98	38	111

4.1. Objedinjavanje pouzdanosti u vremenu

U cilju dobijanja generalizovanih ocena pouzdanosti urađena je analiza: alfa koeficijenata, Loevingerinog koeficijenta, Guttmanov koeficijent pouzdanosti dve polovine, ocena međuposmatračke pouzdanosti ukupnih skorova i ocena test-retest pouzdanosti ukupnih skorova za skale HAMD i MADRS.

4.1.1. Interna konzistentnost za HAMD skalu

4.1.1.1. Analiza alfa koeficijenata za HAMD skalu

Koeficijent alfa za HAMD skalu ekstrahovan je iz 24 longitudinalne studije (0,36% od ukupno 6590 prikupljenih i pregledanih studija) u kojima je korišćena Hamiltonova skala za procenu depresivnosti (Prilog 2). Zbirne karakteristike studija uključenih u meta-analizu prikazane su u Tabeli 3.

Tabela 3. Zbirne karakteristike studija sa ocenama alfa koeficijenta za HAMD skalu

Karakteristike studija	alfa koeficijent	
	Prva procena	Poslednja procena
Ukupan broj ispitanika	5682	4555
Veličina uzorka, medijana (opseg)	115 (18-1629)	
Period praćenja (dani), medijana (opseg)		62 (17-365)
Depresija kao primarna dijagnoza, n (%)	13 (54%)	
Kriterijum za uključenje u studiju u vezi sa težinom depresije, n (%)	9 (37,5%)	
Starost (godine), AS (SD)	43,7 (9,6)	
HAMD ukupni skorovi, objedinjena vrednost	20,4	10,6
HAMD SD, objedinjena vrednost	4,5	6,1

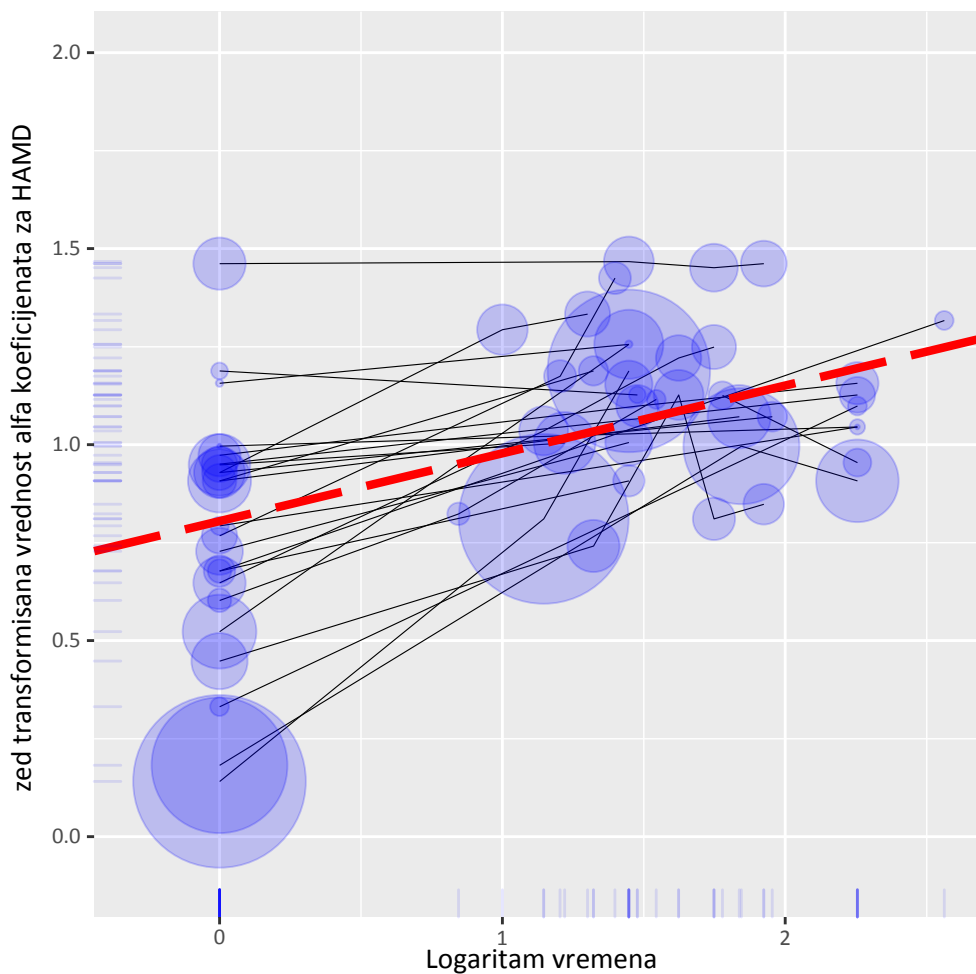
Veličina uzorka na početku studija bila je u opsegu 18 do 1629 ispitanika. I u ovoj meta-analizi veličina uzorka je manja na kraju studija nego na početku, ukupan broj ispitanika na kraju studija manji je za oko 20%. Prema mestu gde su ocenjivani ispitanici tokom istraživanja: 25% je bilo hospitalizovano, 62,5% je bilo ambulantno a 12,5% je bilo i hospitalizovano i ambulantno. Starost ispitanika kretala se u opsegu 30,5-69 godina. Ispitanici su češće bili (66,7%) van engleskog govornog područja.

Trajanje studija kretalo se u opsegu od 17 do 365 dana. Depresija kao primarna dijagnoza bila je zastupljena u više od 50% studija. U 22 studije korišćena je originalna verzija HAMD skale (20 sa 17 i 2 sa 21 stavkom), dok je u 2 studije korišćena neoriginalna verzija HAMD skale i to sa 24 stavke. Postojanje kriterijuma za uključanje u studiju u vezi sa težinom depresije saopšteno je u 9 od 24 studije. Kao kriterijum u 8 studija korišćena je HAMD skala (skor za uključanje 14 i 18), dok je u jednoj studiji kao kriterijum upotrebljena MADRS skala (skor 22 i viši). Prema tipu studije, češće su bile psihometrijske (71%). Samo u 9 studija saopšteno je ko su bili ocenjivači i u 6 od 9 studija to su bili psihijatri ili klinički psiholozi.

Studije su najčešće imale po dve procene depresivnosti primenom HAMD skale (63%), pri čemu se broj procena unutar pojedinačne studije kretao od 2 do 5.

Objedinjene aritmetičke sredine ukupnih skorova za HAMD skalu na poslednjoj proceni su za skoro 50% manje nego na prvoj proceni. Varijabilitet je bio veći na poslednjim merenjima za skoro 35%.

Objedinjena regresiona linija svih studija uključenih u meta-analizu dobijena je iz zed transformisanih alfa koeficijenata za HAMD skalu primenom RVE metoda. Na Grafikonu 1 se vidi da vrednost zed transformisanih alfa koeficijenata za HAMD skalu raste tokom vremena. Postoje statistički značajne promene u vremenu zed transformisanih vrednosti alfa koeficijenata za HAMD skalu ($b=0,173$; $p<0,001$).



Grafikon 1. Dijagram rasturanja sa individualnim linijama zed transformisanih alfa koeficijenata za HAMD skalu u odnosu na logaritam vremena procena depresivnosti unutar studija. Svaka puna linija predstavlja jednu studiju, a markeri predstavljaju zed transformisane alfa koeficijente pojedinačnih procena tokom studije. Težinski koeficijent veličine markera je kvadratni koren iz recipročne varijanse zed transformisanih alfa koeficijenata za HAMD skalu. Objedinjena regresiona linija prikazana je zadebljanom isprekidanom linijom.

Analiza homogenosti pokazala je visok stepen heterogenosti zed transformisanih alfa koeficijenata za HAMD skalu: $\tau^2=0,150$; $I^2=96\%$, pa se pristupilo analizi moderatora primenom meta-regresije.

Table 4. Univarijantni i multivarijantni meta-regresioni modeli za zed transformisanim alfa koeficijentima za HAMD skalu kao zavisnom varijablom

Nezavisna varijabla	Univarijantna meta-regresija		Multivarijantna meta-regresija	
	b	p	b	p
Tip studije (metodološka ili klinička)	-0,063	0,395		
Primarna dijagnoza ispitanika (depresija ili drugi poremećaji)	0,079	0,320		
Kriterijum za uključenje u studiju u vezi sa težinom depresije	-0,133	0,097		
Starost ispitanika (godine)	<0,001	0,897		
Logaritam vremena procena depresivnosti	0,196	<0,001	0,061	0,062
HAMD prosečan skor	-0,040	<0,001	-0,023	0,010
HAMD standardna devijacija	0,156	0,002	0,098	0,004

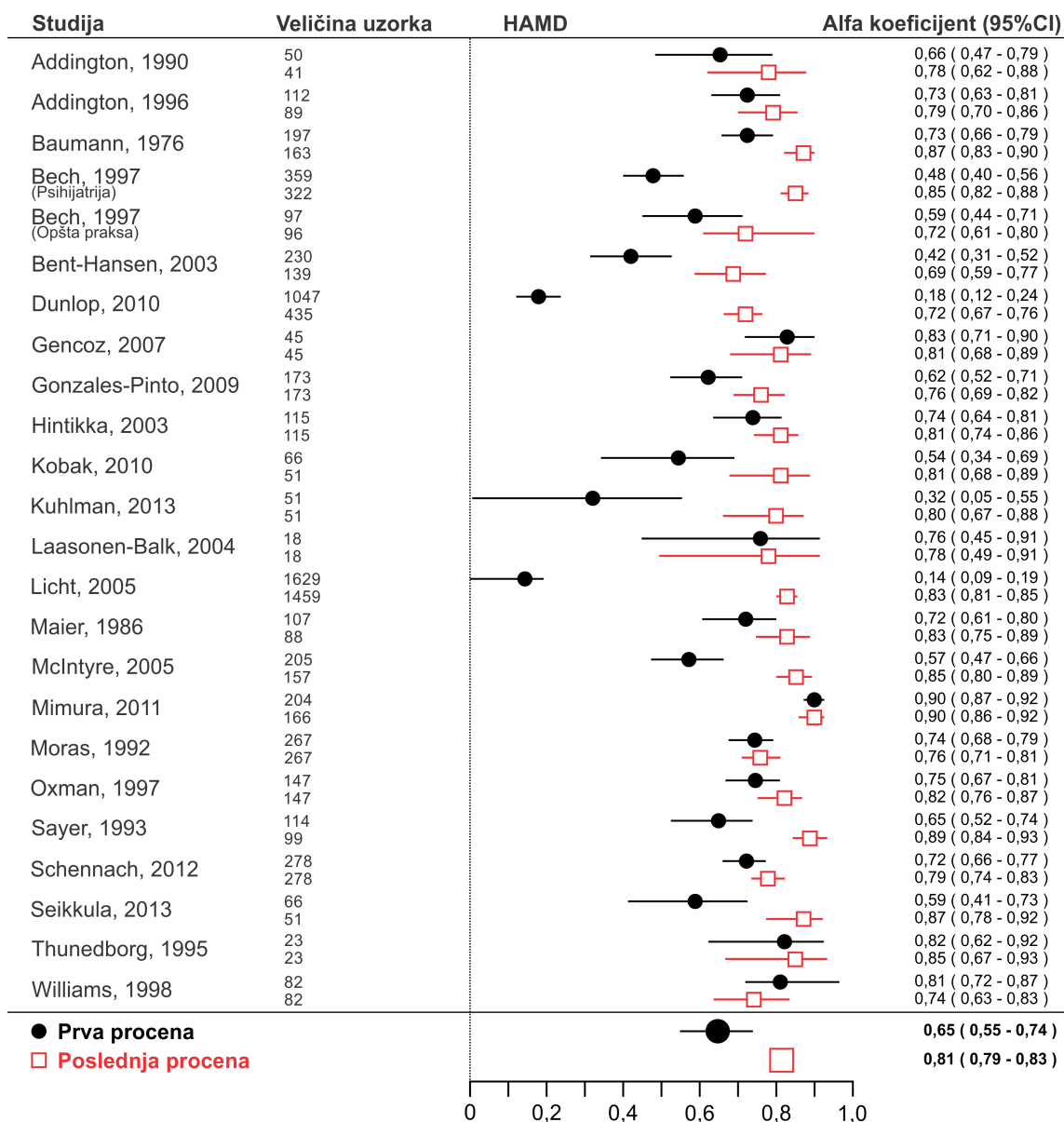
Meta-regersiona analiza izvedena je sa zed transformisanim alfa koeficijentima za HAMD skalu kao zavisnom varijablom. U univarijantnim meta-regresionim modelima, aritmetička sredina skorova na HAMD skali negativno je povezana sa vrednostima zed transformisanih alfa koeficijenata za HAMD skalu. Suprotno od toga, logaritam vremena procena depresivnosti i standardne devijacije skorova na HAMD skali su pozitivno povezane sa vrednostima zed transformisanih alfa koeficijenata za HAMD skalu. Shodno tome, kasnije ponovljene procene depresivnosti i viši varijabilitet skorova na HAMD skali povezani su sa višim vrednostima zed transformisanih alfa koeficijenata za HAMD skalu. U multivarijantnom meta-regresionom modelu, statistički značajni prediktori vrednosti zed transformisanih alfa koeficijenata za HAMD skalu su smanjenje prosečnog

skora na HAMD skali ($p=0,010$) i povećanje standardne devijacije za HAMD skalu ($p=0,004$) (Tabela 4).

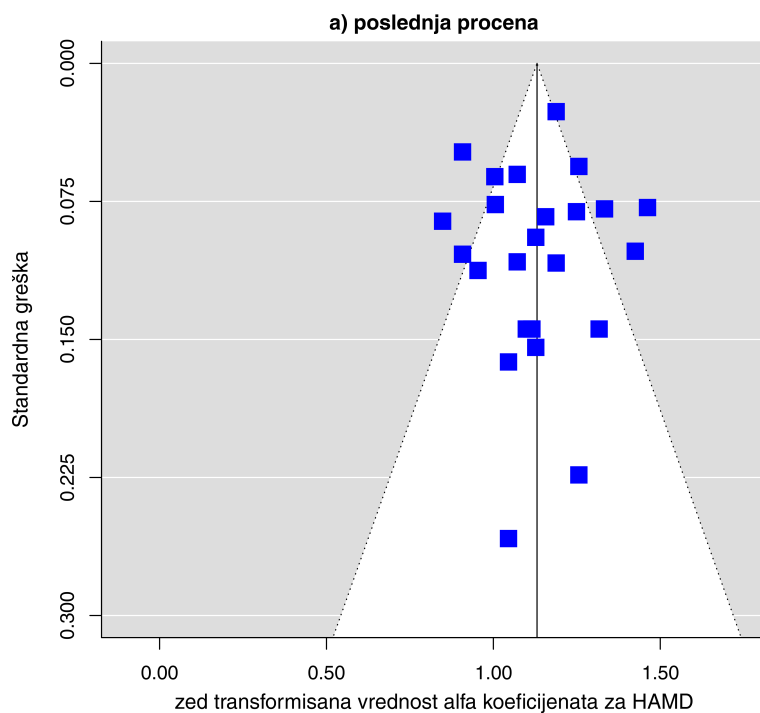
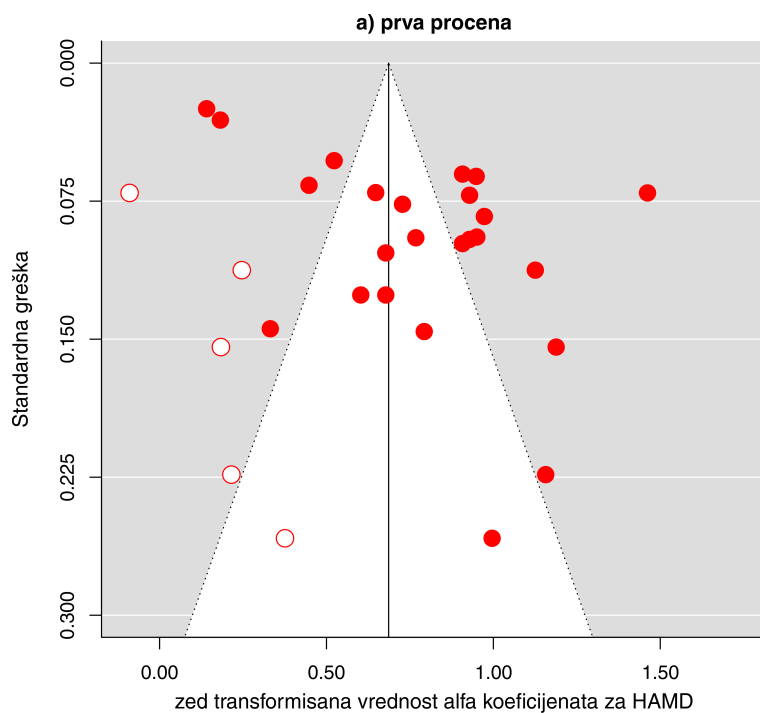
Na Grafikonu 2 prikazan je forest plot alfa koeficijenta za HAMD skalu, sa pripadajućim 95% intervalima poverenja iz pojedinačnih studija, na prvoj i poslednjoj proceni unutar svake studije, respektabilno. U poređenju sa početnim merenjima, vrednosti alfa koeficijenta za HAMD skalu više su na poslednjoj proceni kod 22 (92%) studije uključene u meta-analizu. Objedinjena vrednost alfa koeficijenta za HAMD skalu iz modela slučajnih efekata viša je na poslednjoj proceni. Objedinjene vrednosti alfa koeficijenta za HAMD skalu su sledeće: 0,655 (95% CI 0,548 – 0,741) na prvoj i 0,784 (95% CI 0,787 – 0,833) na poslednjoj proceni.

Publikaciona pristrasnost nije nađena niti primenom levkastog dijagrama niti primenom Eggerovog testa. Levkasti dijagrami na prvom i poslednjem merenju izgledaju dovoljno simetrično (Grafikon 3), a Eggerov test nije bio statistički značajan ($p>0,05$) za prvu i poslednju procenu.

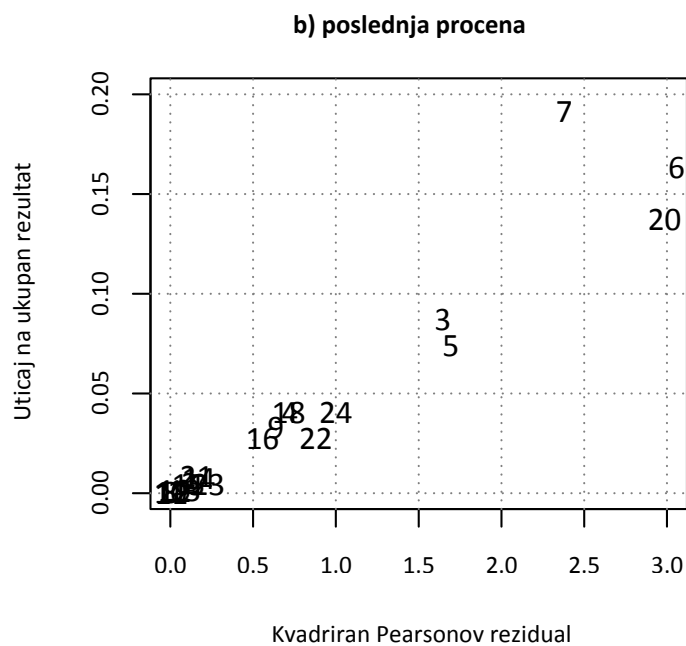
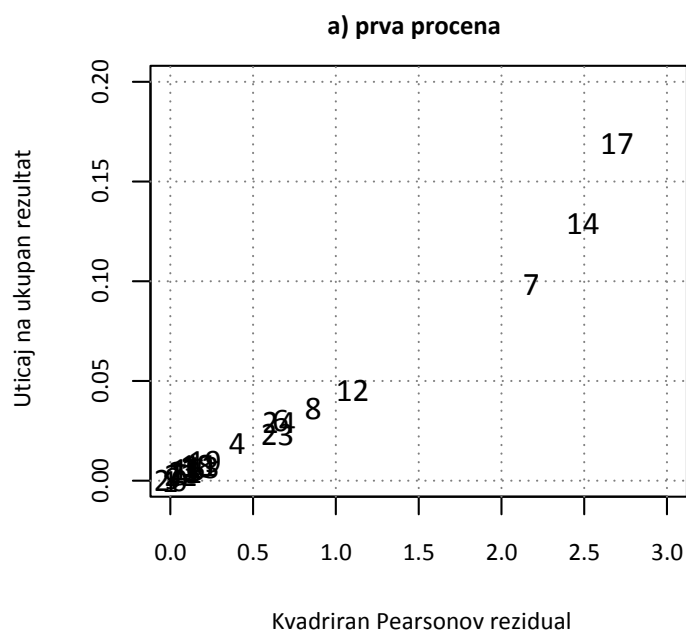
Na prvoj proceni Q test heterogenosti bio je statistički značajan ($p<0,001$), $\tau^2=0,165$ i $I^2=97\%$, što ukazuje na visok stepen heterogenosti. Na poslednjoj proceni Q test heterogenosti bio je statistički značajan ($p<0,001$), $\tau^2=0,019$ i $I^2=76\%$, što ukazuje na visok stepen heterogenosti. Izgled Baujatovih dijagrama na prvoj i poslednjoj proceni, takođe ukazuje na heterogenost studija.



Grafikon 2. Forest plot sa vrednostima alfa koeficijenata za HAMD skalu i pripadajućim 95% intervalima poverenja iz pojedinačnih studija na prvoj i poslednjoj proceni.



Grafikon 3. Levkasti dijagrami (sa ukloniti i popuniti metodom dodavanja nedostajućih studija) zed transformisanih alfa koeficijena za HAMD skalu na prvoj (a) i poslednjoj (b) proceni. Eggerov test: na prvoj proceni $p=0,058$; na poslednjoj proceni $p=0,922$.



Grafikon 4. Baujatov dijagram zed transformisanih alfa koeficijenata za HAMD skalu na prvoj i poslednjoj proceni.

4.1.1.2. Analiza Loevingerinog koeficijenta za HAMD skalu

Loevingerin koeficijent za HAMD skalu ekstrahovan je iz 10 longitudinalnih studija (0,15% od ukupno 6590 prikupljenih studija) u kojima je korišćena Hamiltonova skala za procenu depresivnosti (Prilog 3). U svim studijama uključenim u meta-analizu korišćena je originalna verzija HAMD skale. U studiji Bech i sar. (2004) korišćena je verzija HAMD skale sa 24 stavke ali je Loevingerin koeficijent saopšten za prvih 17 ajtema. Studija Bech i sar. (2009) koristila je podatke iz dve druge studije Fava i sar. (2006) i Martiny i sar. (2004). Zbirne karakteristike studija uključenih u meta-analizu prikazane su u Tabeli 5.

Tabela 5. Zbirne karakteristike studija sa saopštenim Loevingerinim koeficijentom za HAMD skalu

Karakteristike studija	Loevingerin koeficijent	
	Prva procena	Poslednja procena
Ukupan broj ispitanika	3698	3346
Veličina uzorka, medijana (opseg)	210 (23-1629)	
Period praćenja (dani), medijana (opseg)		42 (27-84)
Depresija kao primarna dijagnoza, n (%)	9 (90%)	
Kriterijum za uključenje u studiju u vezi sa težinom depresije, n (%)	8 (80%)	
Starost (godine), AS (SD)	45,2 (7,1)	
HAMD ukupni skorovi, objedinjena vrednost	22,4	13,3
HAMD SD, objedinjena vrednost	3,7	6,3

AS: aritmetička sredina; SD: standardna devijacija

Veličina uzorka na početku studija je u opsegu 23 do 1629 ispitanika. I u ovoj meta-analizi veličina uzorka je manja na kraju studija nego na početku, ukupan broj ispitanika smanjio se za oko 10%. Ispitanici su najčešće bili ambulantni (90%). Prosečna starost ispitanika po studijama kretala se od 39 do 59 godina. Prema jezičkom području

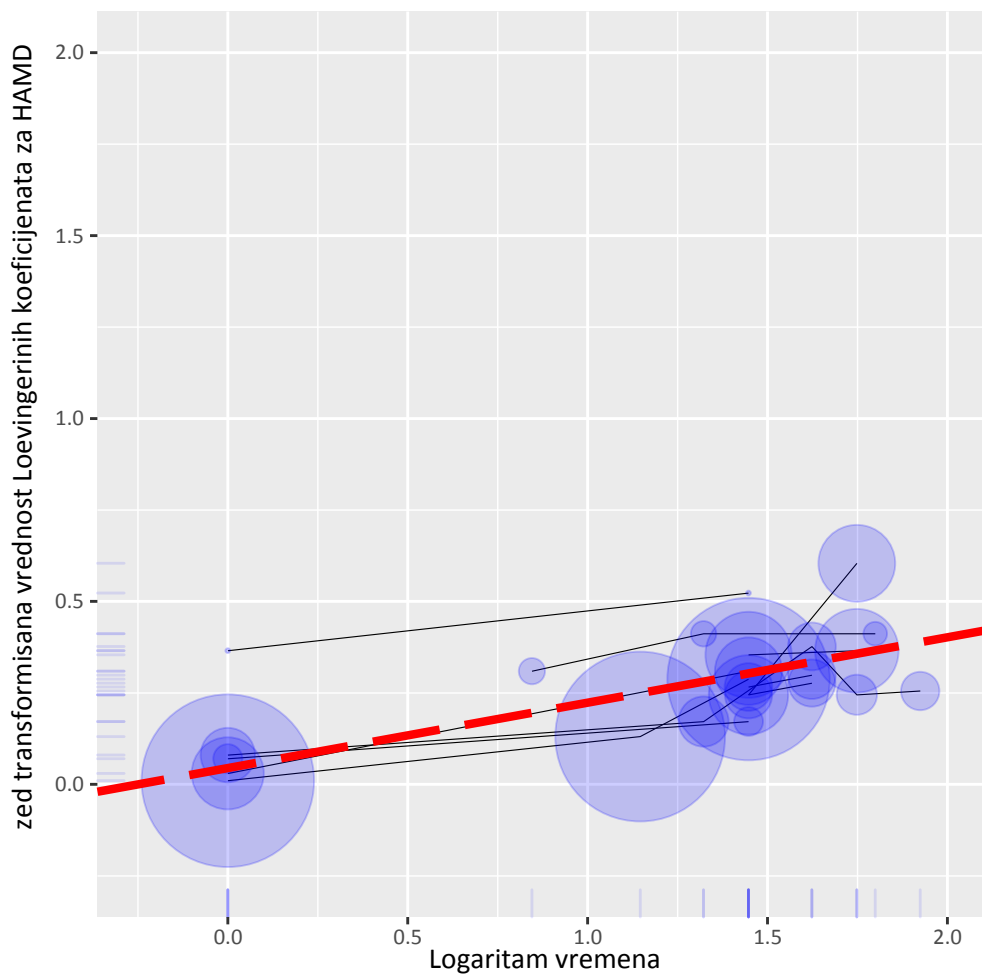
ispitanika, Engleski jezik bio je zastupljen u 2, dok su ostali jezici bili zastupljeni u drugih 8 studija.

Trajanje studija kretalo se u opsegu od 28 do 84 dana. Depresija kao primarna dijagnoza bila je zastupljena u 90% studija. Postojanje kriterijuma za uključanje u studiju u vezi sa težinom depresije saopšteno je u 8 studija, od toga je za kriterijum korišćena HAMD skala u 6, dok se u 2 studije koristila MADRS skala. Skor za uključanje kretao se u opsegu 13 do 25. Prema tipu studija, psihometrijskih je bilo 7, dok su 3 su bile kliničke. Samo u jednoj studiji (Bent-Hansen i sar, 2003) bilo je saopšteno ko je ocenjivao pacijente, i da su to su bili psihijatri.

Studije su najčešće imale po dve procene depresivnosti primenom MADRS skale, pri čemu se broj procena unutar pojedinačne studije kretao od 2 do 5.

Objedinjene aritmetičke sredine ukupnih skorova za HAMD skalu na poslednjoj proceni su za 40% manje nego na prvoj proceni. Varijabilitet je bio veći na poslednjim merenjima za više od 70%.

Objedinjena regresiona linija svih studija uključenih u meta-analizu dobijena je iz zed transformisanog Loevingerinog koeficijenta za HAMD skalu koristeći RVE metod. Na Grafikonu 5 se vidi da vrednost zed transformisanih Loevingerinih koeficijenta za HAMD skalu raste tokom vremena i taj porast je statistički značajan ($b=0,179$; $p=0,002$).



Grafikon 5. Dijagram rasturanja sa individualnim linijama zēd transformisanih Loevingerinih koeficijenta za HAMD skalu u odnosu na logaritam vremena procena depresivnosti unutar studija. Svaka puna linija predstavlja jednu studiju, a markeri predstavljaju zēd transformisane Loevingerine koeficijente pojedinačnih procena tokom studije. Težinski koeficijent veličine markera je kvadratni koren iz recipročne varijanse zēd transformisanih Loevingerinih koeficijenata za HAMD. Objedinjena regresiona linija prikazana je zadebljanom isprekidanom linijom.

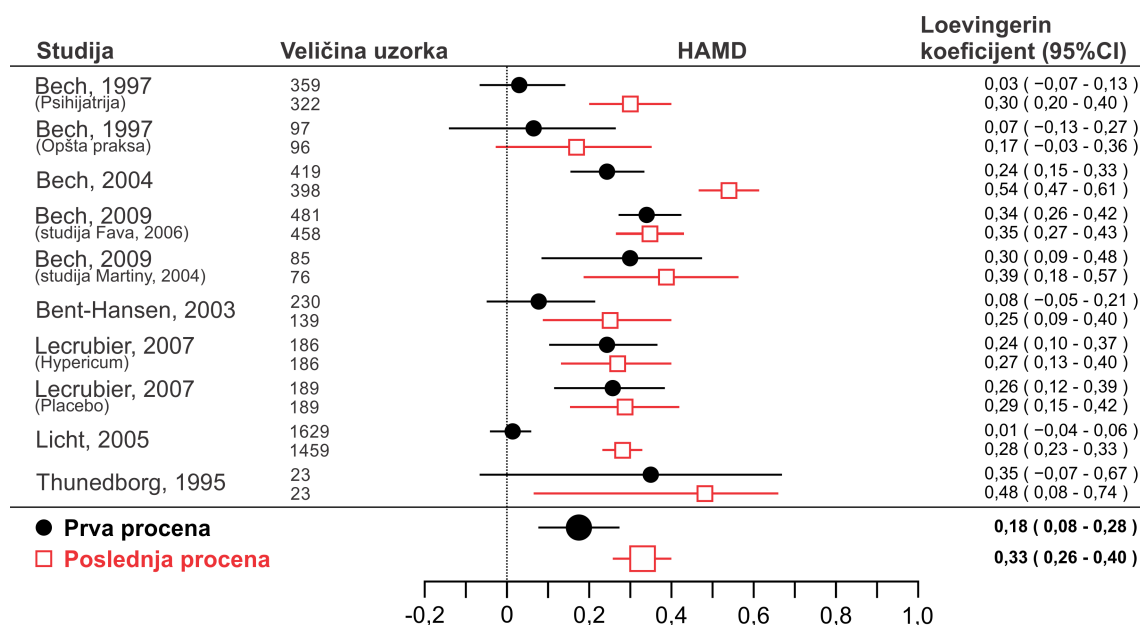
Analiza homogenosti pokazala je visok stepen heterogenosti zed transformisanog Loevingerinog koeficijenta za HAMD skalu: $I^2=66\%$; $\tau^2=0,007$, pa se pristupilo analizi moderatora primenom meta-regresije.

Tabela 6. Univarijantni meta-regresioni modeli za zed transformisanim Loevingerinim koeficijentima za HAMD skalu kao zavisnom varijablom

Nezavisna varijabla	Univarijantna meta-regresija	
	b	p
Tip studije (metodološka ili klinička)	-0,036	0,743
Primarna dijagnoza ispitanika (depresija ili drugi poremećaji)	-	-
Kriterijum za uključenje u studiju u vezi sa težinom depresije	0,155	0,105
Starost ispitanika (godine)	-	-
Logaritam vremena procena depresivnosti	0,148	0,004
HAMD prosečan skor	-	-
HAMD standardna devijacija	-	-

Meta-regersiona analiza izvedena je sa zed transformisanim Loevingerinim koeficijentom za HAMD skalu kao zavisnom varijablom. U univarijantnim meta-regresionim modelima jedini statistički značajan prediktor porasta zed transformisanog Loevingerinog koeficijenta za HAMD skalu je logaritam vremena (Tabela 6). Univarijantni regresioni modeli sa varijablama primarna dijagnoza, starost ispitanika, HAMD prosečan skor i HAMD standardna devijacija zbog numeričkih ograničenja za RVE metod nisu izvedeni.

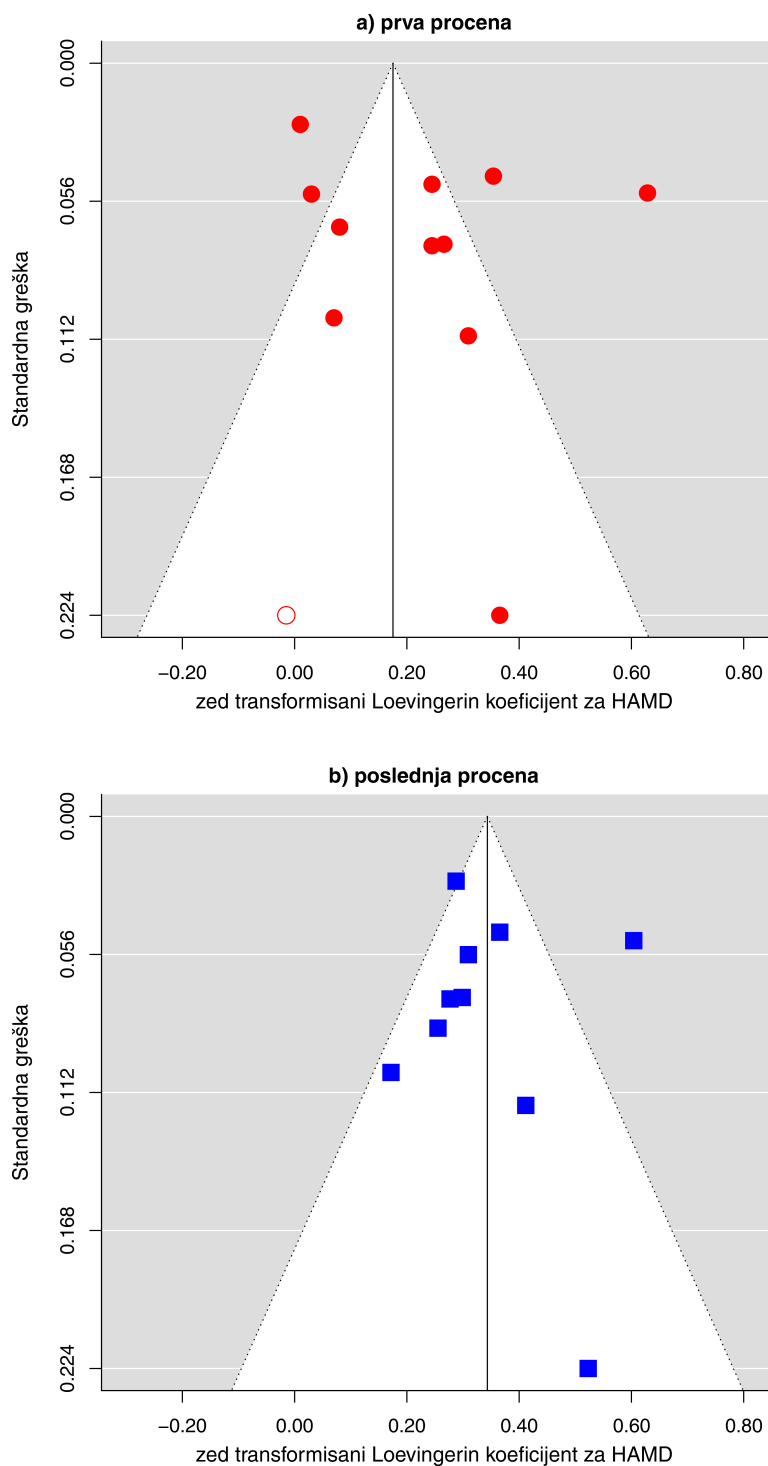
U poređenju sa početnim merenjima, vrednosti zed transformisanih Loevingerinih koeficijenata za HAMD skalu više su na poslednjoj proceni kod svih studija uključenih u meta-analizu. Objedinjena vrednost zed transformisanih Loevingerinih koeficijenata za HAMD skalu iz modela slučajnih efekata su sledeće: 0,181 (95% CI 0,080 – 0,277) na prvom i 0,331 (95% CI 0,255 – 0,402) na poslednjem merenju.



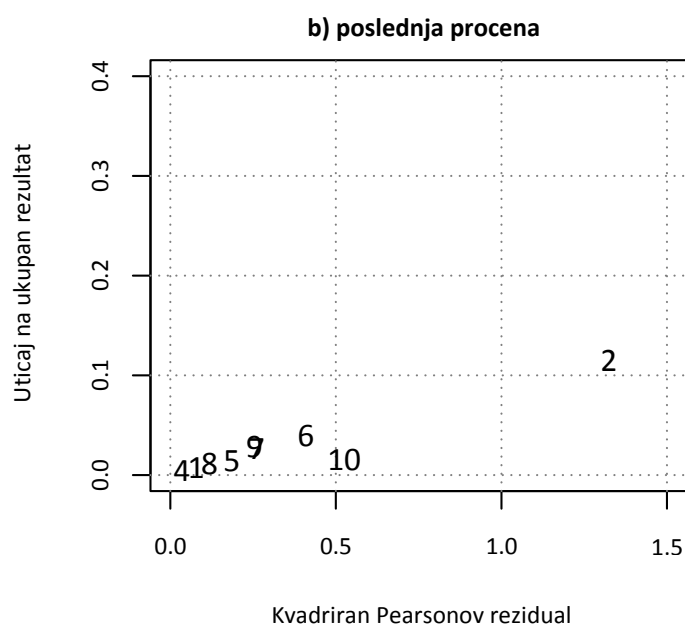
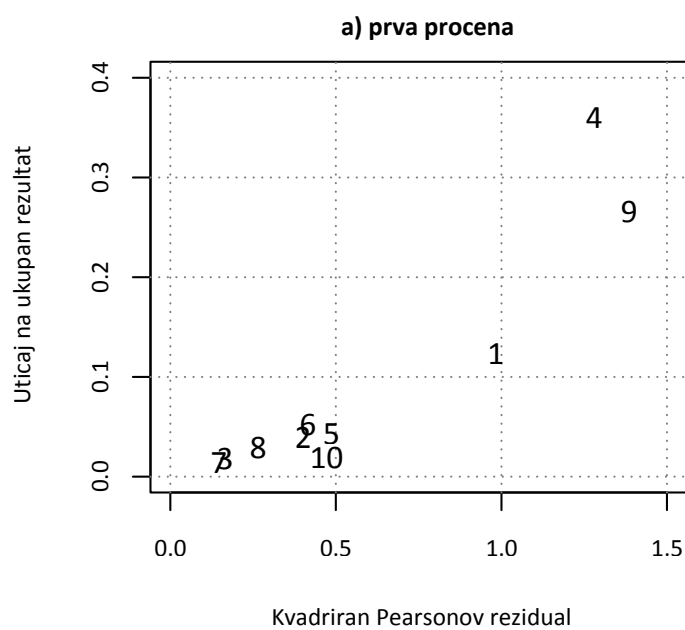
Grafikon 6. Forest plot sa vrednostima Loevingerinovog koeficijenta za HAMD skalu i pripadajućim 95% intervalima poverenja iz pojedinačnih studija na prvoj i poslednjoj proceni.

Publikaciona pristrasnost nije nađena niti primenom levkastog dijagrama niti primenom Eggerovog testa. Levkasti dijagrami na prvoj i poslednjoj proceni izgledaju dovoljno simetrično (grafikon 7), a Eggerov test nije bio statistički značajan ($p > 0,05$) za prvu i poslednju procenu.

Na prvoj proceni Q test heterogenosti bio je statistički značajan ($p < 0,001$), $\tau^2 = 0,021$ i $I^2 = 87\%$, što ukazuje na visok stepen heterogenosti. Na poslednjoj proceni Q test heterogenosti bio je statistički značajan ($p < 0,001$), $\tau^2 = 0,011$ i $I^2 = 76\%$ što ukazuje na visok stepen heterogenosti. Izgled Baujatovih dijagrama na prvoj i poslednjoj proceni, takođe ukazuje na heterogenost studija.



Grafikon 7. Levkasti dijagrami (sa ukloniti i popuniti metodom dodavanja nedostajućih studija) zed transformisanih Loevingerinih koeficijenata za HAMD skalu na prvoj (a) i poslednjoj (b) proceni. Eggerov test: na prvoj proceni $p=0,334$; na poslednjoj proceni $p=0,920$.



Grafikon 8. Baujatov dijagram zed transformisanih Loevingerinih koeficijenata za HAMD skalu na prvoj i poslednjoj proceni.

4.1.1.3. Pouzdanost dve polovine

Guttmanov koeficijent pouzdanosti dve polovine za HAMD skalu

Samo studija Gencoz i sar. (2007) imala je saopštene vrednosti Guttmanovog koeficijenta pouzdanosti dve polovine za HAMD skalu u ponovljenim merenjima. Studija je psihometrijska, pacijenti nemaju depresiju kao primarnu dijagnozu. Guttmanov koeficijent pouzdanosti dve polovine saopšten je za dva merenja (na početku studije i u periodu praćenja) sa razlikom od 30 dana. Koeficijent je izračunat na 45 pacijenata. Na prvoj proceni vrednost Guttmanovog koeficijenta pouzdanosti dve polovine iznosi 0,82 a na drugoj 0,85.

4.1.2. Interna konzistentnost za MADRS skalu

4.1.2.1. Analiza alfa koeficijenata za MADRS skalu

Koeficijent alfa za MADRS skalu ekstrahovan je iz 7 longitudinalnih studija (0,34% od ukupno 2051 prikupljenih i pregledanih studija) u kojima je korišćena Montgomeri-Ašbergova skala za procenu depresivnosti (Prilog 2). Zbirne karakteristike studija uključenih u meta-analizu prikazane su u Tabeli 7.

Tabela 7. Zbirne karakteristike studija sa ocenama alfa koeficijenta za MADRS skalu

Karakteristike studija	alfa koeficijent	
	Prva procena	Poslednja procena
Ukupan broj ispitanika	2665	2502
Veličina uzorka, medijana (opseg)	283 (63-1000)	
Period praćenja (dani), medijana (opseg)		180 (27-364)
Depresija kao primarna dijagnoza, n (%)	6 (86%)	
Kriterijum za uključenje u studiju u vezi sa težinom depresije, n (%)	4 (57%)	
Starost (godine), AS (SD)	51,5 (14,4)	
HAMD ukupni skorovi, objedinjena vrednost	24,5	9,1
HAMD SD, objedinjena vrednost	5,3	7,1

AS: aritmetička sredina; SD: standardna devijacija

Veličina uzorka na početku studija bila je u opsegu 63 do 1000 ispitanika. I u ovoj meta-analizi veličina uzorka je manja na kraju studija nego na početku, ukupan broj ispitanika na poslednjoj proceni manji je za oko 6%. Prema mestu gde su ocenjivani ispitanici tokom istraživanja: 71% je bilo ambulantno a 14% je bilo i hospitalizovano i ambulantno. Za jednu studiju nije bilo saopšteno da li su pacijenti ambulantni ili

hospitalizovani. Starost ispitanika kretala se u opsegu 40-74,5 godina. Ispitanici su češće bili (86%) van engleskog govornog područja.

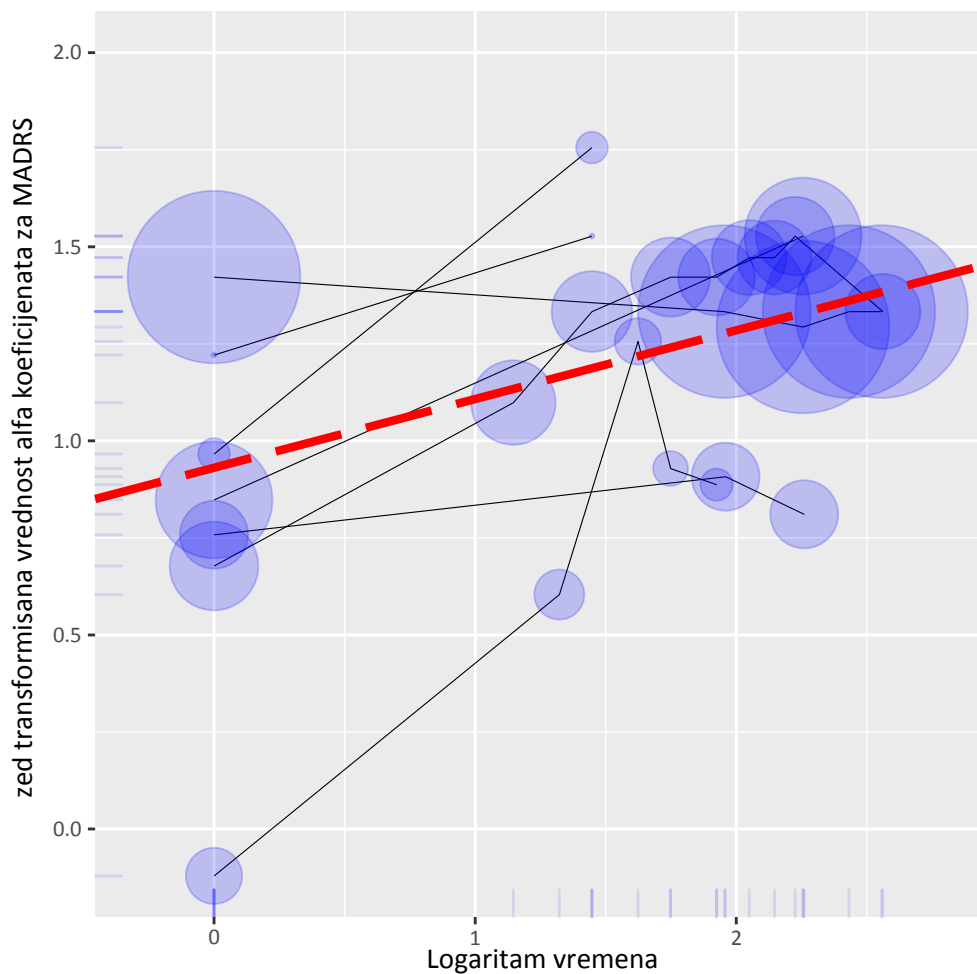
Trajanje studija kretalo se u opsegu od 28 do 364 dana. Depresija kao primarna dijagnoza bila je zastupljena u više od 2/3 studija. U svim studijama korišćena je originalna verzija MADRS skale sa 10 ajtema. Postojanje kriterijuma za uključenje u studiju u vezi sa težinom depresije saopšteno je u 7 od 7 studija. Kao kriterijum upotrebljavana je MADRS skala (skor 20-22 i viši). Prema tipu studije, češće su bile psihometrijske (71%). Ocenjivači su u 5 od 6 studija bili psihijatri ili klinički psiholozi.

Studije su najčešće imale po dve procene depresivnosti primenom MADRS skale (43%), pri čemu se broj procena unutar pojedinačne studije kretao od 2 do 9.

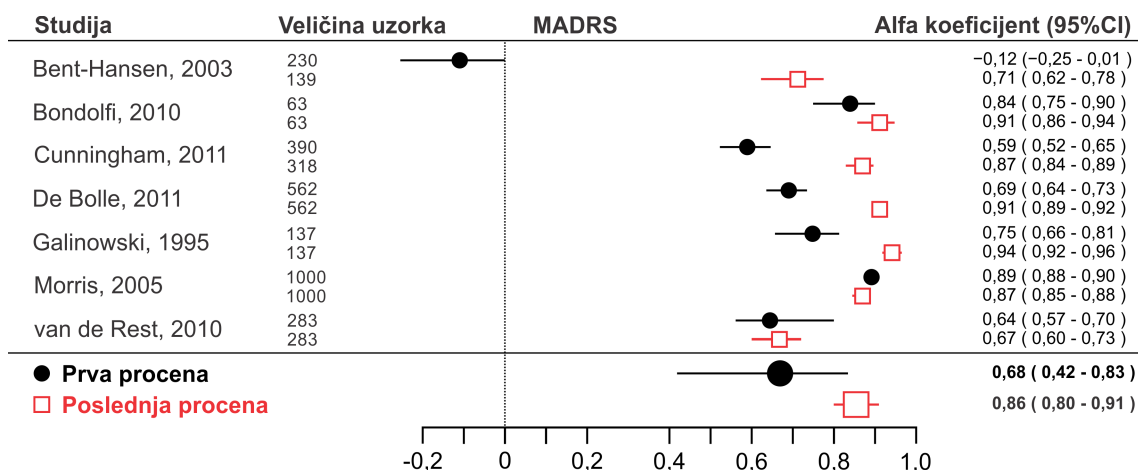
Objedinjene aritmetičke sredine ukupnih skorova za MADRS skalu na poslednjoj proceni su za skoro 50% manje nego na prvoj proceni. Varijabilitet je bio veći na poslednjim merenjima za skoro 35%.

Objedinjena regresiona linija svih studija uključenih u meta-analizu dobijena je iz zed transformisanih alfa koeficijenata za MADRS skalu primenom RVE metoda. Na Grafikonu 9 se vidi da vrednost zed transformisanih alfa koeficijenata za MADRS skalu raste tokom vremena. Nisu dobijene statistički značajne promene u vremenu za vrednosti zed transformisanih alfa koeficijenata za MADRS skalu ($b=0,177$; $p=0,065$), što je verovatno posledica malog uzorka.

Zbog numeričkih ograničenja RVE metoda i nepovoljnog odnosa broja studija i potencijalnih prediktora nije bilo moguće uraditi meta-regresionu analizu sa zed transformisanim vrednostima alfa koeficijenata za MADRS skalu kao zavisnom varijablom.



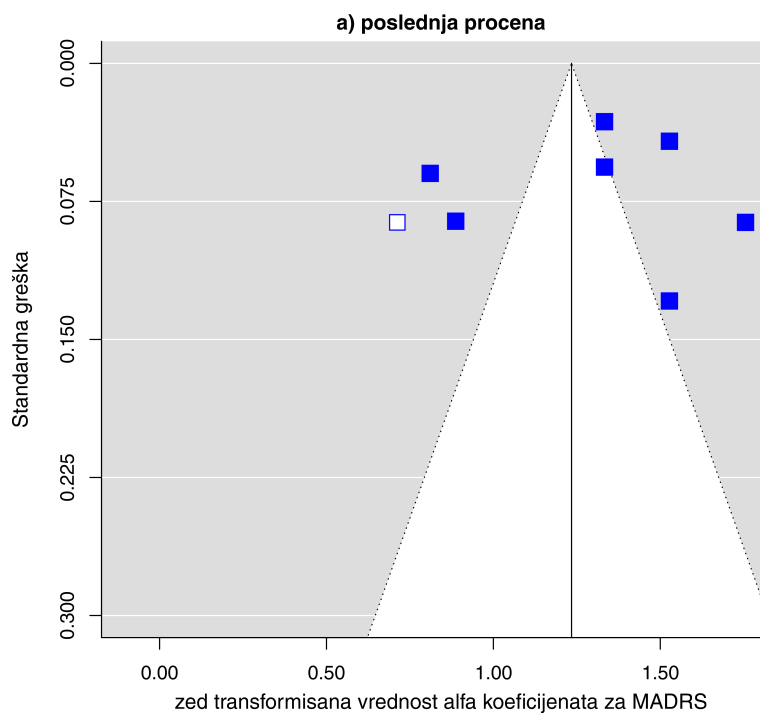
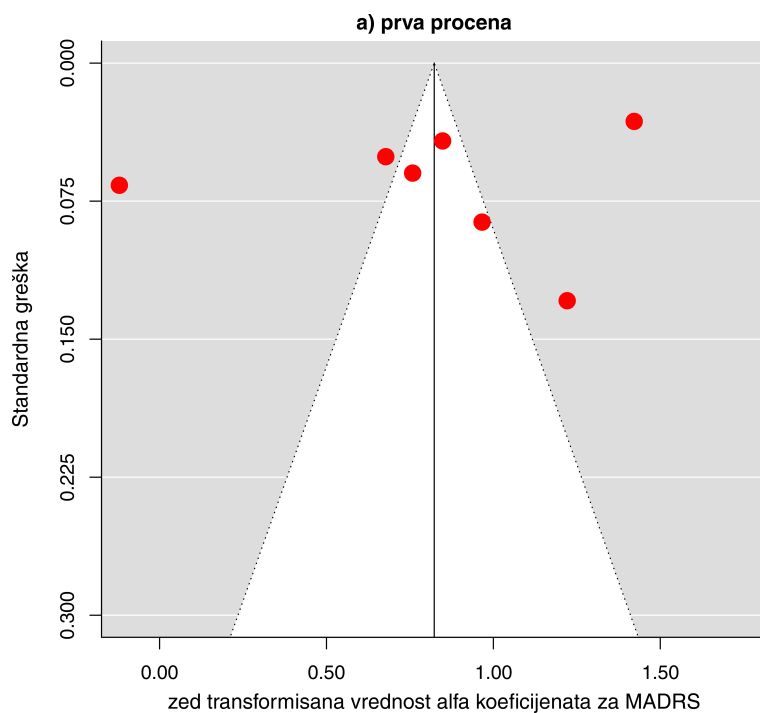
Grafikon 9. Dijagram rasturanja sa individualnim linijama zed transformisanih alfa koeficijenata za MADRS skalu u odnosu na logaritam vremena procena depresivnosti unutar studija. Svaka puna linija predstavlja jednu studiju, a markeri predstavljaju zed transformisane alfa koeficijente pojedinačnih procena tokom studije. Težinski koeficijent veličine markera je kvadratni koren iz recipročne varijanse zed transformisanih alfa koeficijenata za MADRS skalu. Objedinjena regresiona linija prikazana je zadebljanom isprekidanom linijom.



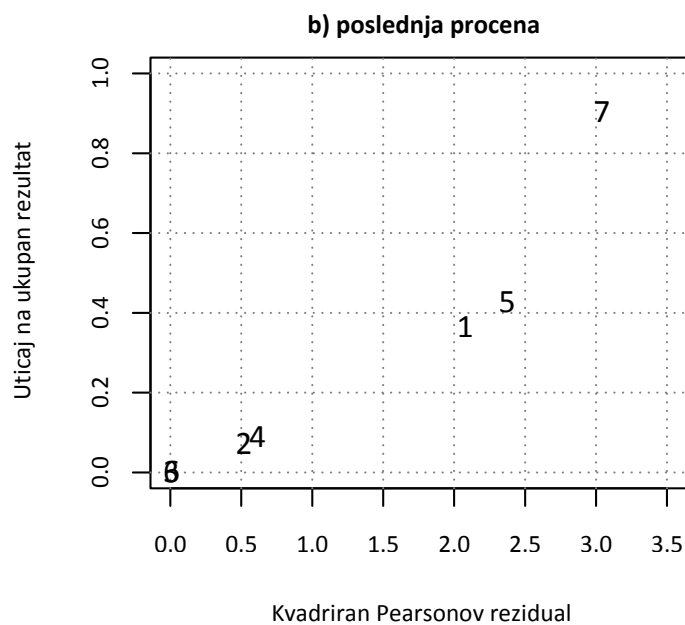
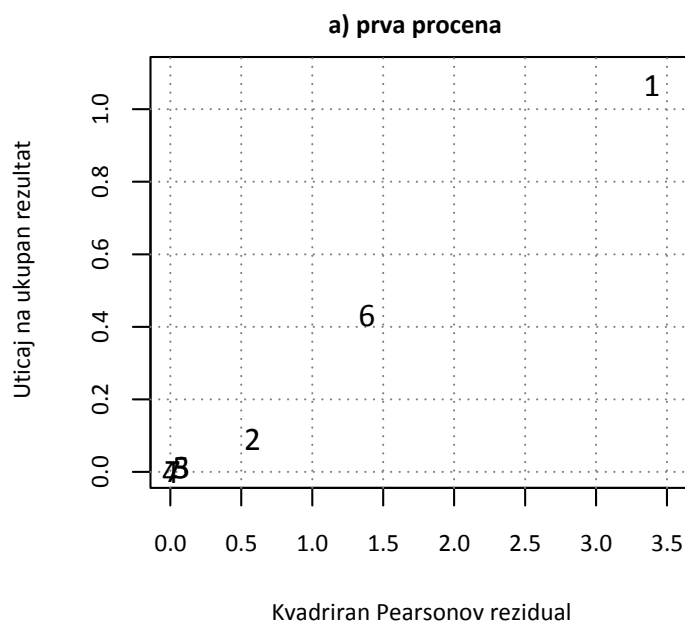
Grafikon 10. Forest plot sa vrednostima alfa koeficijenta za MADRS skalu i pripadajućim 95% intervalima poverenja iz pojedinačnih studija na prvoj i poslednjoj proceni.

Publikaciona pristrasnost nije nađena niti primenom levkastog dijagrama niti primenom Eggerovog testa. Levkasti dijagrami na prvom i poslednjem merenju izgledaju dovoljno simetrično (Grafikon 11), a Eggerov test nije bio statistički značajan ($p > 0,05$).

Na prvoj proceni Q test heterogenosti bio je statistički značajan ($p < 0,001$), $\tau^2 = 0,258$ i $I^2 = 99\%$, što ukazuje na visok stepen heterogenosti. Na poslednjoj proceni Q test heterogenosti bio je statistički značajan ($p < 0,001$), $\tau^2 = 0,077$ i $I^2 = 96\%$ što ukazuje na visok stepen heterogenosti. Izgled Baujatovih dijagrama na prvoj i poslednjoj proceni, takođe ukazuje na heterogenost studija.



Grafikon 11. Levkasti dijagrami (sa ukloniti i popuniti metodom dodavanja nedostajućih studija) zed transformisanih alfa koeficijena za MADRS skalu na prvoj (a) i poslednjoj (b) proceni. Eggerov test: na prvoj proceni $p=0,844$; na poslednjoj proceni $p=0,693$.



Grafikon 12. Baujatov dijagram zed transformisanih alfa koeficijenata za MADRS skalu na prvoj i poslednjoj proceni.

4.1.2.2. Analiza Loevingerinog koeficijenta za MADRS skalu

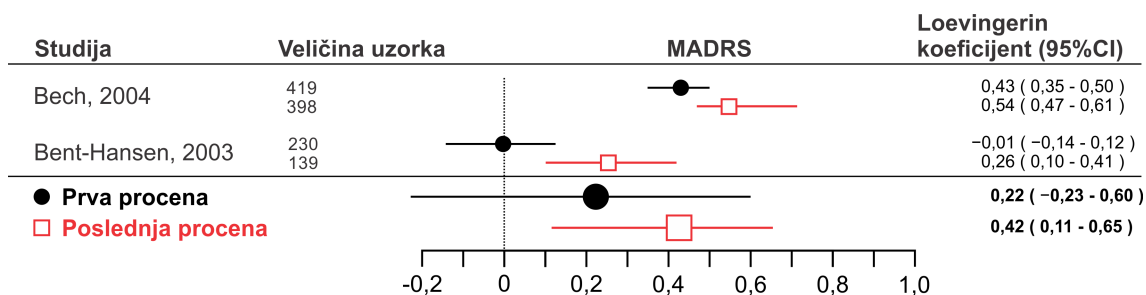
Loevingerin koeficijent za MADRS skalu ekstrahovan je iz 2 longitudinalne studije (0,1% od ukupno 2051 prikupljenih studija) u kojima je korišćena Montgomery-Ašbergova skala za procenu depresivnosti (Prilog 3).

Studija Bech i sar. (2004) je klinička, sa ambulantnim pacijentima čija je primarna dijagnoza bila depresija. U okviru studije bile su dve procene MADRS skale. Studija je trajala 56 dana, započelo je 416 ispitanika, od kojih je 395 završilo studiju. Prosečna starost ispitanika bila je 40,1 godinu. Ispitanici su bili sa engleskog govornog područja.

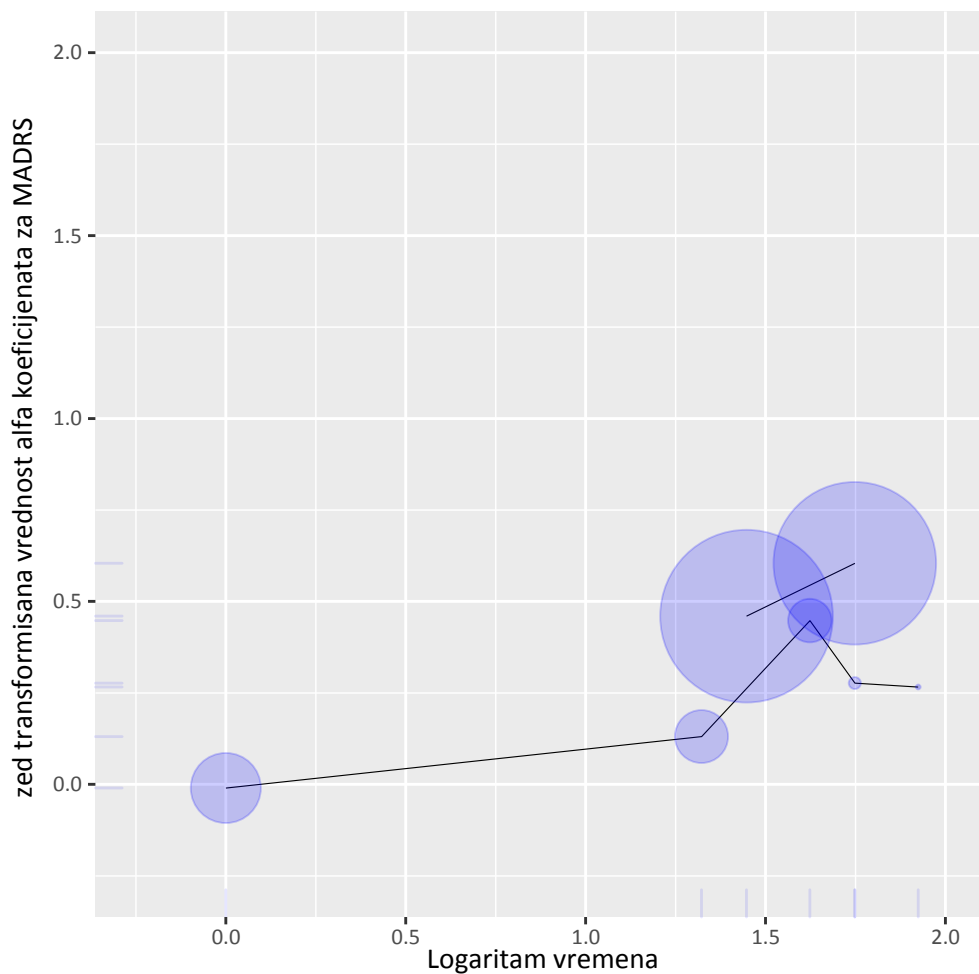
Druga uključena studija, Bent-Hansen i sar. (2003) je psihometrijska studija, sa ambulantnim pacijentima čija je primarna dijagnoza bila depresija. U okviru studije bilo je 5 procena MADRS skale. Na početku studije bilo je 230 ispitanika, od kojih je 139 završilo studiju koja je trajala 84 dana. Ispitanici su bili van engleskog govornog područja.

U obe studije kao kriterijuma za uključenje u studiju u vezi sa težinom depresije bo je 22 i viši skor na MADRS skali.

Objedinjene vrednosti na prvoj proceni za MADRS ukupan skor iznosi 22,5, a za standardnu devijaciju MADRS iznosi 4,2.



Grafikon 13. Forest plot sa vrednostima Loevingerinog koeficijenta za MADRS skalu i pripadajućim 95% intervalima poverenja iz pojedinačnih studija na prvoj i poslednjoj proceni.



Grafikon 14. Dijagram rasturanja sa individualnim linijama zed transformisanih Loevingerinih koeficijenata za MADRS skalu u odnosu na logaritam vremena procena depresivnosti unutar studija. Svaka puna linija predstavlja jednu studiju, a markeri predstavljaju zed transformisane Loevingerine koeficijente pojedinačnih procena tokom studije. Težinski koeficijent veličine markera je kvadratni koren iz recipročne varijanse zed transformisanih Loevingerinih koeficijenata za MADRS skalu.

U poređenju sa početnim merenjima, vrednosti zed transformisanih Loevingerinih koeficijenata za MADRS skalu više su na poslednjoj proceni kod obe studije uključene u meta-analizu. Objedinjena vrednost zed transformisanih Loevingerinih koeficijenata za MADRS iz modela slučajnih efekata su sledeće: 0,223 (95% CI -0,229 – 0,596) na prvom i 0,416 (95% CI 0,111 – 0,649) na poslednjem merenju.

4.1.3. Međuposmatračka pouzdanost

4.1.3.1. Analiza ocena međuposmatračke pouzdanosti ukupnih skorova za HAMD skalu

Iz 19 longitudinalnih studija ekstrahovane su 24 ocene međuposmatračke pouzdanosti u ponovljenim merenjima. Od tog broja 22 koeficijenta (91,7%) saopštena su kao precizna ocena a u 2 studije saopšteno je kao neprecizna ocena u vidu intervala ili jednostrane granične vrednosti.

Intraklasni koeficijent korelacije je najčešće saopšten (63,2%), dok je kappa koeficijent saopšten samo u jednoj studiji.

Tabela 8. Struktura ocena međuposmatračke pouzdanosti za HAMD skalu prema načinu izračunavanja

Način izračunavanja ocene međuposmatračke pouzdanosti	Način saopštavanja ocene					
	Ocena u vidu intervala					
	Precizna ocena		ili jednostrane granične vrednosti		Svega	
	Broj	%	Broj	%	Broj	%
ICC	11	64,7	1	50,0	12	63,2
Kappa	1	5,9	-	-	1	5,3
Drugi načini	2	11,8	1	50,0	3	15,8
Nespecificovano	3	17,6	-	-	3	15,8
Ukupno	17	100,0	2	100,0	19	100,0

Karakteristike studija u kojima su saopštene ocene međuposmatračke pouzdanosti za HAMD skalu prikazane su u Prilogu 4.

4.1.3.1.1. Analiza intraklasnog koeficijenta korelacije za HAMD skalu

Intraklasni koeficijent korelacije za HAMD skalu ekstrahovan je iz 8 longitudinalnih studija (0,12% od ukupno 6590 prikupljenih i pregledanih studija) u kojima je korišćena Hamiltonova skala za procenu depresivnosti. Zbirne karakteristike studija uključenih u meta-analizu prikazane su u Tabeli 9.

Tabela 9. Zbirne karakteristike studija ICC za HAMD skalu

Karakteristike studija	ICC	
	Prva procena	Poslednja procena
Ukupan broj ispitanika	556	-
Veličina uzorka, medijana (opseg)	27 (7-81)	
Period praćenja (dani), medijana (opseg)		42 (21-112)
Depresija kao primarna dijagnoza, n (%)	8 (100%)	
Kriterijum za uključanje u studiju u vezi sa težinom depresije, n (%)	6 (75%)	
Starost (godine), AS (SD)	30,4 (12,8)	
HAMD ukupni skorovi, objedinjena vrednost	21,9	13,7
HAMD SD, objedinjena vrednost	6,6	8,6

AS: aritmetička sredina; SD: standardna devijacija

Veličina uzorka na kojoj je izračunavan ICC za HAMD skalu na početku studije kretala se u opsegu od 7 do 81 ispitanik. Ukupan broj uključenih ispitanika po studijama nije saopšten u 6 studija na poslednjoj proceni, pa nije bilo moguće izračunati tu vrednost. Ispitanici su podjednako praćeni ambulantno i bolnički (po 3 studije), dok za dve studije nije saopšteno koji su ispitanici uključeni. Prosečna starost ispitanika bila je saopštena samo u 3 studije i kretala se u opsegu 15,7-38,8 godina. Prema jezičkom području ispitanika, Engleski jezik bio je zastupljen u 87,5% uključenih studija.

Trajanje studija kretalo se u opsegu od 21 do 112 dana. Depresija kao primarna dijagnoza bila je zastupljena kod svih studija (100%). U 7 studija uključenih u meta-

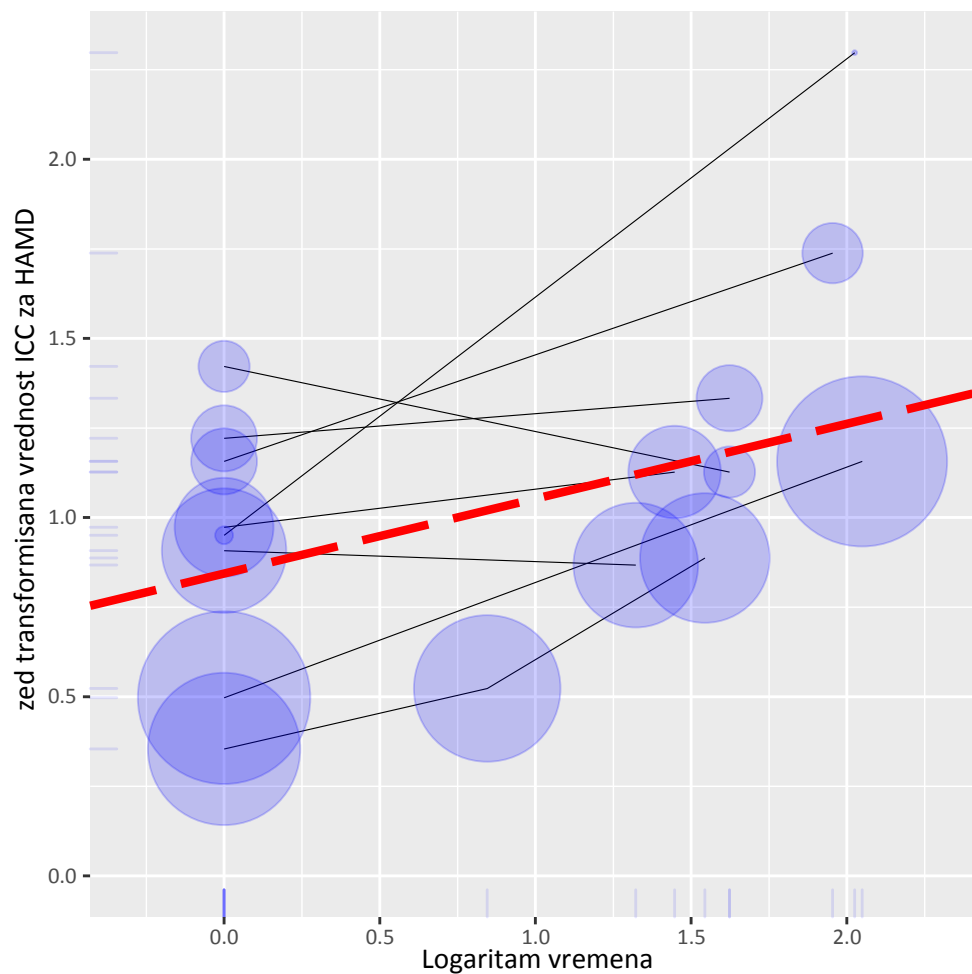
analizu korišćene su originalne verzije HAMD skale (u 5 studija sa 17 ajtema), dok je u studiji Williams i sar. (2008) korišćena HAMD-GRID verzija skale. Postojanje kriterijuma za uključenje u studiju u vezi sa težinom depresije saopšteno je u 6 od 8 studija. Kao kriterijum u 4 studije korišćen je HAMD skala (skor za uključenje 16 i 17), dok je u po jednoj studiji kriterijum bila BDI skala (skor 20 i viši) ili RDS (skor 7 i viši). Prema tipu uključenih studija, po 50% su bile zastupljene psihometrijske ili kliničke studije.

Ocenjivači su u 50% slučajeva bili psihijatri ili klinički psiholozi. Trening za ocenjivače u proceni HAMD skale postojao je u 5 studija, dok su u samo 3 studije ocenjivači imali ranije iskustvo u proceni HAMD skale. Prema načinu na koji je sprovedena procena intervjua ispitanika, u 2 studije rađen je zajednički intervju, po jedna studija koristila je video ili audio zapis, dok je najveći broj uključenih studija (50%) koristio individualni intervju kao način sprovođenja procene ispitanika.

Studije su najčešće imale po dve procene depresivnosti primenom HAMD skale (87,5%), jedino je studija Kobak i sar (2010) imala 3 procene.

Objedinjene aritmetičke sredine ukupnih skorova za HAMD skali na poslednjoj proceni su za 37% manje nego na prvoj proceni. Varijabilitet je bio veći na poslednjim merenjima za 30%.

Objedinjena regresiona linija svih studija uključenih u pojedinačnu meta-analizu dobijena je iz zed transformisanog ICC za HAMD skali koristeći RVE metod. Na Grafikonu 15 se vidi da vrednost zed transformisanog ICC za HAMD skalu raste tokom vremena. Postoje statistički značajne promene u vremenu zed transformisanih vrednosti ICC za HAMD skalu ($b=0,209$; $p=0,025$).



Grafikon 15. Dijagram rasturanja sa individualnim linijama zed transformisanih ICC za HAMD skalu u odnosu na logaritam vremena procena depresivnosti unutar studija. Svaka puna linija predstavlja jednu studiju, a markeri predstavljaju zed transformisane ICC pojedinačnih procena tokom studije. Težinski koeficijent veličine markera je kvadratni koren iz recipročne varijanse zed transformisanih ICC za HAMD skalu. Objedinjena regresiona linija prikazana je zadebljanom isprekidanom linijom.

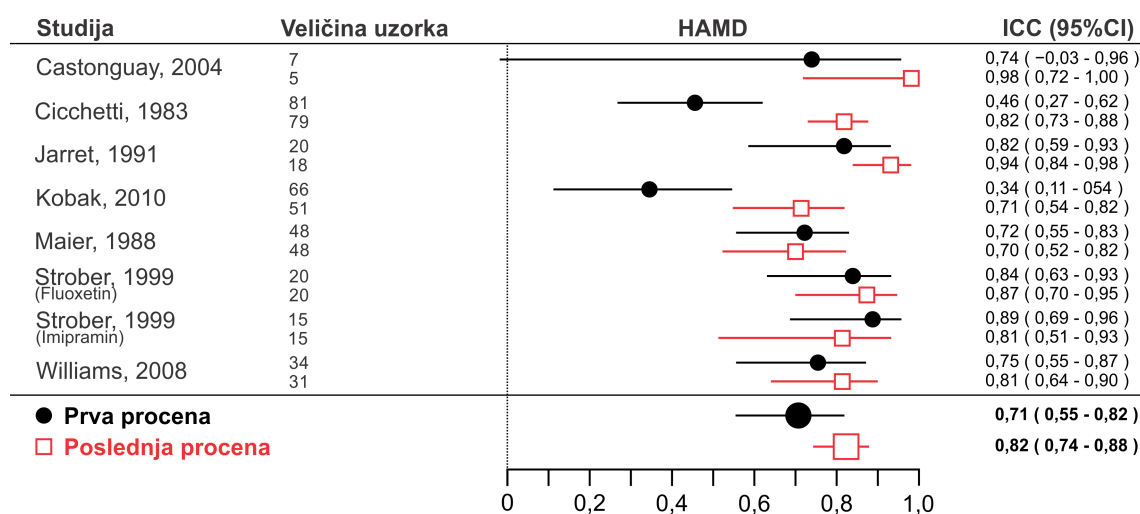
Analiza homogenosti pokazala je visok stepen heterogenosti zed transformisanog ICC za HAMD: $I^2=77\%$; $\tau^2=0,116$, pa se pristupilo analizi moderatora primenom meta-regresije.

Tabela 10. Univarijantni meta-regresioni modeli za zed transformisanim ICC za HAMD skalu kao zavisnom varijablom

Nezavisna varijabla	Univarijantna meta-regresija	
	b	p
Tip studije (metodološka ili klinička)	0,533	0,005
Primarna dijagnoza ispitanika (depresija ili drugi poremećaji)	-	-
Kriterijum za uključenje u studiju u vezi sa težinom depresije	0,117	0,592
Starost ispitanika (godine)	-0,009	0,266
Logaritam vremena procena depresivnosti	0,202	0,032
HAMD prosečan skor	-0,022	0,175
HAMD standardna devijacija	-	-

Meta-regersiona analiza izvedena je sa zed transformisanim ICC za HAMD skalu kao zavisnom varijablom. U univarijantnim meta-regresionim modelima statistički značajni prediktori porasta zed transformisanog ICC za HAMD skalu u vremenu su tip studije i logaritam vremena procena depresivnosti (Tabela 10). Shodno tome kliničke studije u odnosu na psihometrijske i kasnije ponovljene procene depresivnosti povezane su sa višim vrednostima zed transformisanih ICC za HAMD. Univarijantni regresioni modeli sa varijablama primarna dijagnoza i HAMD standardna devijacija zbog numeričkih ograničenja za RVE metod nisu izvedeni.

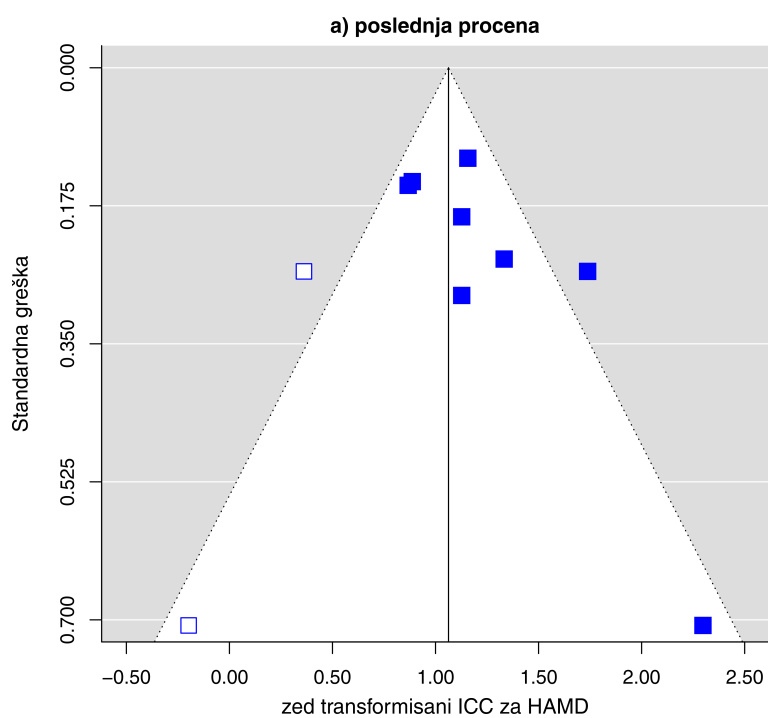
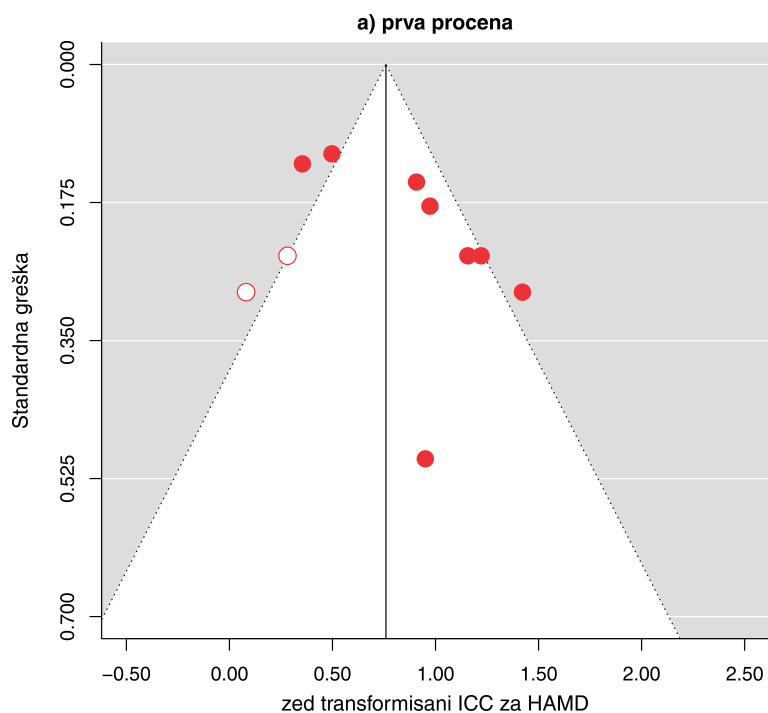
Na grafikonu 16 prikazan je forest plot ICC za HAMD skalu, sa pripadajućim 95% intervalima poverenja iz pojedinačnih studija, na prvoj i poslednjoj proceni unutar svake studije, respektabilno. U poređenju sa početnim merenjima, vrednosti ICC za HAMD skalu više su na poslednjoj proceni kod 6 (75%) studija uključenih u meta-analizu. Objedinjena vrednost ICC za HAMD skalu iz modela slučajnih efekata je viša na poslednjoj proceni. Objedinjene vrednosti ICC za HAMD skalu su sledeće: 0,708 (95% CI 0,546 – 0,819) na prvom i 0,819 (95% CI 0,740 – 0,875) na poslednjem merenju.



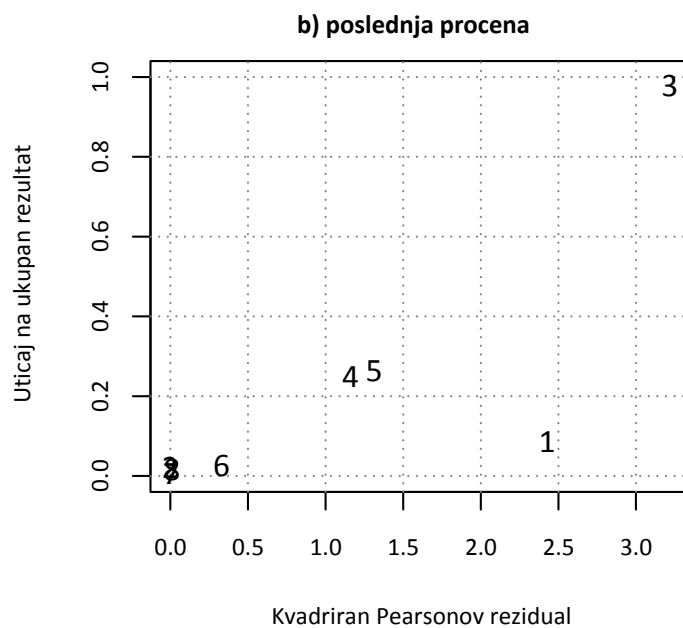
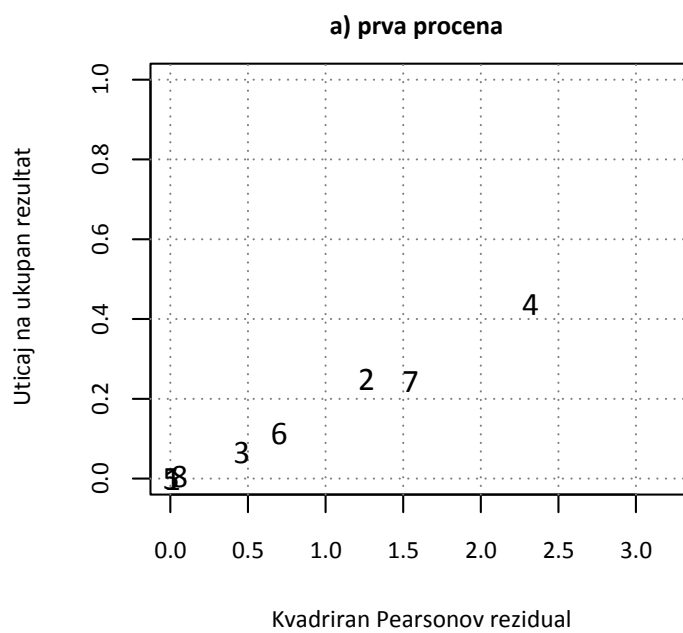
Grafikon 16. Forest plot sa vrednostima ICC za HAMD skalu i pripadajućim 95% intervalima poverenja iz pojedinačnih studija na prvoj i poslednjoj proceni.

Publikaciona pristrasnost nađena je na prvoj proceni primenom levkastog dijagrama (grafikon 17 pod a) i primenom Eggerovog testa ($p=0,014$). I na poslednjoj proceni levkasti dijagram izgleda asimetrično (grafikon 17 pod b), a Eggerov test bio je statistički značajan ($p=0,020$), što ukazuje na publikacionu pristrasnost.

Na prvoj proceni Q test heterogenosti bio je statistički značajan ($p<0,001$), $\tau^2=0,105$ i $I^2=76\%$, što ukazuje na visok stepen heterogenosti. Na poslednjoj proceni Q test heterogenosti bio je statistički značajan ($p=0,039$), $\tau^2=0,040$ i $I^2=53\%$ što ukazuje na srednji stepen heterogenosti. Izgled Baujatovih dijagrama na prvoj i poslednjoj proceni, takođe ukazuje na heterogenost studija.



Grafikon 17. Levkasti dijagrami (sa ukloniti i popuniti metodom dodavanja nedostajućih studija) zed transformisanih ICC za HAMD skalu na prvoj (a) i poslednjoj (b) proceni. Eggerov test: na prvoj proceni $p=0,014$; na poslednjoj proceni $p=0,020$.



Grafikon 18. Baujatov dijagram zed transformisanih ICC za HAMD skalu na prvoj i poslednjoj proceni.

4.1.3.1.2. Analiza kappa koeficijenta za HAMD skalu

Samo studija Peyre i sar. (1989) imala je saopštene vrednosti kappa za HAMD u ponovljenim merenjima. Studija je psihometrijska, pacijenti su sa depresijom kao primarnom dijagnozom. Autori nisu naveli na koliko pacijenata je rađen kappa koeficijent i nisu naveli vreme ocenjivanja. Na prvoj proceni vrednost kappa koeficijenta pouzdanosti iznosi 0,78 a na drugoj 0,83.

4.1.3.2. *Analiza ocena međuposmatračke pouzdanosti ukupnih skorova za MADRS skalu*

Iz 7 longitudinalnih studija ekstrahovano je 7 ocena međuposmatračke pouzdanosti u ponovljenim merenjima. Od tog broja 2 studije su saopštile ICC, 3 su saopštile kappa koeficijent, dok su u 2 studije bili saopšteni nespecificovani koeficijenti pouzdanosti ili su izračunati na neki drugi način.

Tabela 11. Struktura ocena međuposmatračke pouzdanosti za MADRS skalu prema načinu izračunavanja

Način izračunavanja ocene međuposmatračke pouzdanosti	Način saopštavanja ocene				Svega	
	Precizna ocena		Ocena u vidu intervala ili jednostrane granične vrednosti			
	Broj	%	Broj	%	Broj	%
ICC	2	29	-	-	2	29
Kappa	3	43	-	-	3	43
Drugi načini	1	14	-	-	1	14
Nespecificovano	1	14	-	-	1	14
Ukupno	7	100	-	-	7	100

Karakteristike studija u kojima su saopštene ocene međuposmatračke pouzdanosti za MADRS skalu prikazane su u Prilogu 4.

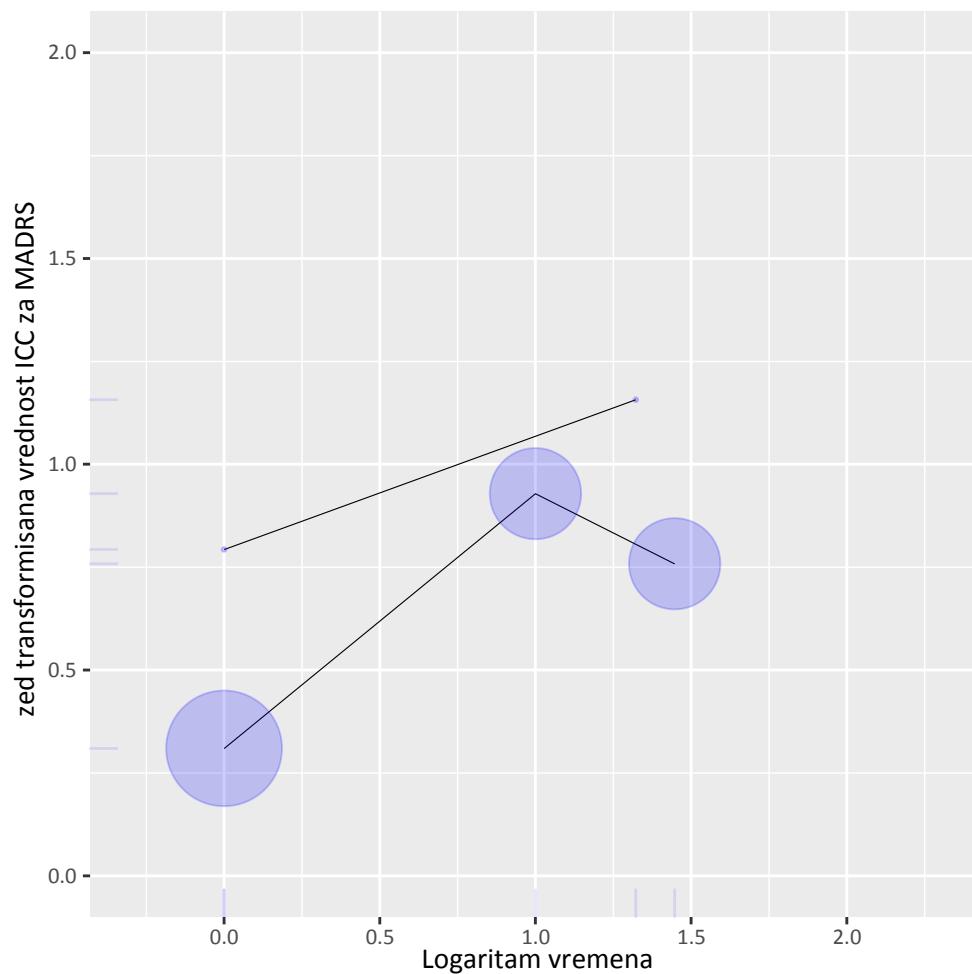
4.1.3.2.1. Analiza intraklasnog koeficijenta korelacije za MADRS skalu

Intraklasni koeficijent korelacije za MADRS skalu ekstrahovan je iz 2 longitudinalne studije (0,1% od ukupno 2051 prikupljenih studija) u kojima je korišćena Montgomeri-Ašbergova skala za procenu depresivnosti.

Studija Maier i sar. (1988) je psihometrijska, sa ambulantnim pacijentima čija je primarna dijagnoza bila depresija. U okviru studije bile su dve procene MADRS skale. Studiju je trajala 21 dan, svi ispitanici (48) su završili studiju. Ispitanici uključeni u studiju bili su van engleskog govornog područja. Nije postojao kriterijum za uključenje u studiju u vezi sa težinom depresije. Ocenjivači su bili psihijatri, postojao je trening za procenu MADRS skale i imali su prethodnog iskustva. Prema načinu na koji je sprovedena procena intervju ispitanika, studija je koristila individualni intervju.

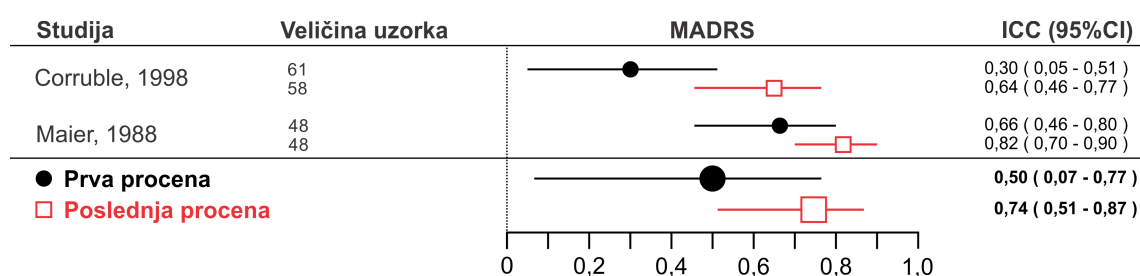
Druga uključena studija, Studija Corruble i sar. (1998) je psihometrijska studija, sa ambulantnim pacijentima čija je primarna dijagnoza bila depresija. U okviru studije bile su tri procene MADRS skale. Studiju je trajala 28 dana. Od 61 uključenog ispitanika, studiju je završilo 58 ispitanika. Prosečna starost ispitanika iznosila je 40,9 godina. Ispitanici uključeni u studiju bili su van engleskog govornog područja. MADRS skor viši od 20 je bio kriterijum za uključenje u studiju u vezi sa težinom depresije. Ocenjivači su bili psihijatri, postojao je trening za procenu MADRS skale i imali su prethodnog iskustva. Prema načinu na koji je sprovedena procena intervju ispitanika, studija je koristila video zapis.

Objedinjene vrednosti za prosečan MADRS ukupan skor na prvoj proceni iznosi 28,7, dok na drugoj proceni iznosi 15,8. Objedinjene vrednosti za MADRS standardne devijacije na prvoj proceni iznosi SD 7,9, dok na drugoj proceni iznosi SD 10,4. Došlo je do smanjenja ukupnih skorova i povećanja standardne devijacije u vremenu.



Grafikon 19. Dijagram rasturanja sa individualnim linijama zed transformisanih ICC za MADRS skalu u odnosu na logaritam vremena procena depresivnosti unutar studija. Svaka puna linija predstavlja jednu studiju, a markeri predstavljaju zed transformisane ICC pojedinačnih procena tokom studije. Težinski koeficijent veličine markera je kvadratni koren iz recipročne varijanse zed transformisanih ICC za MADRS skalu. Objedinjena regresiona linija prikazana je zadebljanom isprekidanom linijom.

Na Grafikonu 20 prikazan je forest plot ICC za MADRS skalu, sa pripadajućim 95% intervalima poverenja iz pojedinačnih studija, na prvoj i poslednjoj proceni unutar svake studije, respektabilno. U poređenju sa početnim merenjima, vrednosti ICC za MADRS skalu više su na poslednjoj proceni kod obe studije uključene u meta-analizu. Objedinjena vrednost ICC za MADRS skalu iz modela slučajnih efekata je viša na poslednjoj proceni. Objedinjene vrednosti ICC za MADRS skalu su sledeće: 0,498 (95% CI 0,082 – 0,770) na prvom i 0,741 (95% CI 0,509 – 0,872) na poslednjem merenju.



Grafikon 20. Forest plot sa vrednostima ICC za MADRS skalu i pripadajućim 95% intervalima poverenja iz pojedinačnih studija na prvoj i poslednjoj proceni.

Zbog numeričkih ograničenja RVE metoda i nepovoljnog odnosa broja studija i potencijalnih prediktora meta-regresija sa zed transformisanim vrednostima alfa koeficijenata za MADRS skalu zavisnom varijablom nije izvođena.

4.1.3.2.2. Analiza kappa koeficijenta za MADRS skalu

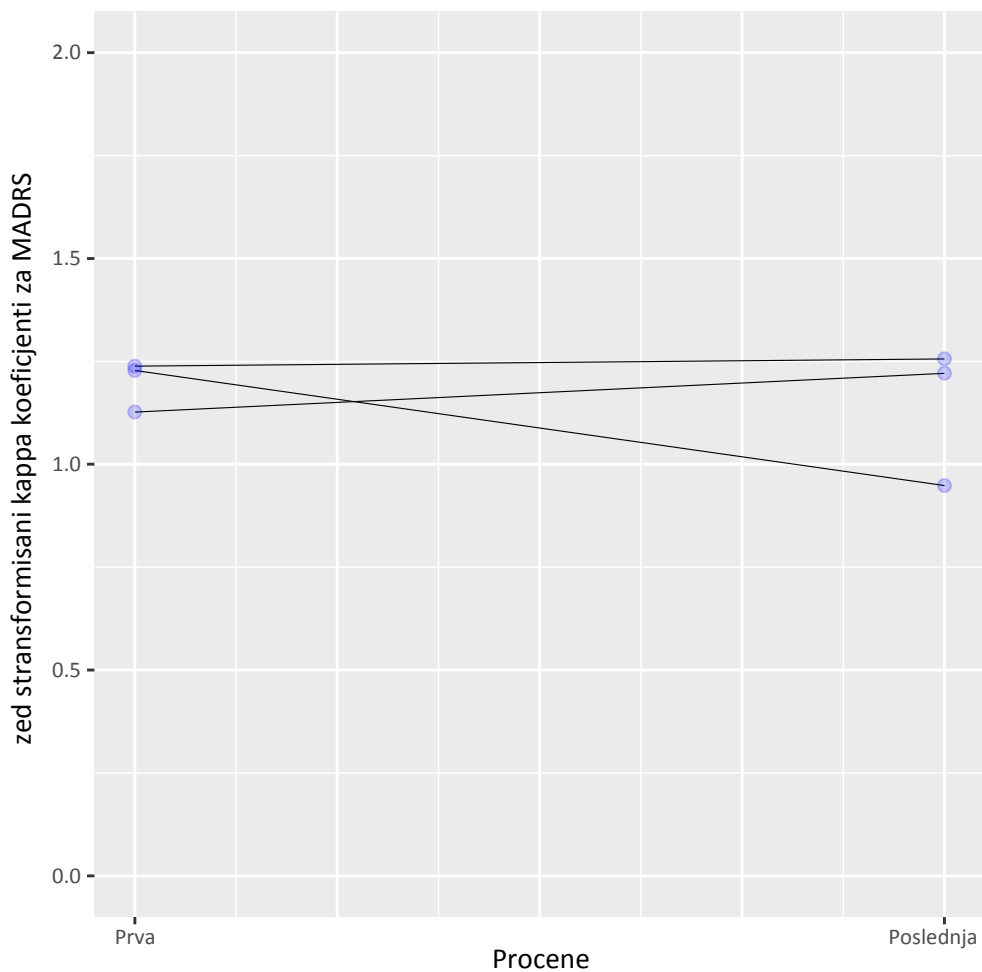
Kappa koeficijent za MADRS skalu ekstrahovan je iz 3 longitudinalne studije (0,15% od ukupno 2051 prikupljenih studija) u kojima je korišćena Montgomeri-Ašbergova skala za procenu depresivnosti.

Studija Bauer i sar. (2009) je klinička studija, sa ambulantnim pacijentima i depresijom kao primarnom dijagnozom. U okviru studije bile su dve procene MADRS skale, na početku i u periodu praćenja. Studiju je započelo 493 a kompletiralo 424 ispitanika, što je smanjenje za 14%. Prosečna starost ispitanika bila je 45,4 godine.

Druga uključena studija El-Khalili (2010) je klinička studija, sa ambulantnim pacijentima i depresijom kao primarnom dijagnozom. U okviru studije bile su dve procene MADRS skale, na početku i u periodu praćenja. Studiju je započelo 446 a kompletiralo 259 ispitanika, što je smanjenje za 42%. Prosečna starost ispitanika bila je 45,5 godine.

Treća studija uključena u meta-analizu je studija Peyre i sar. (1989), koja je opisana u delu za HAMD skalu.

Kriterijuma za uključenje u studiju u vezi sa težinom depresije saopšten je u studijama Bauer i sar. (2009) i El-Khalili (2010) kao ukupan skor na HAMD skali viši ili jednak 22.



Grafikon 21. Dijagram rasturanja sa individualnim linijama zed transformisanih kappa koeficijenata za MADRS skalu na prvoj i poslednjoj proceni. Svaka puna linija predstavlja jednu studiju, a markeri predstavljaju zed transformisane kappa koeficijente pojedinačnih procena tokom studije.

Neponderisane objedinjene vrednosti kapa za MADRS skalu su sledeće: 0,833 (95% CI 0,780 – 0,874) na prvoj i 0,815 (95% CI 0,619 – 0,916) na poslednjoj proceni.

4.1.4. Test-retest pouzdanost

Dodatni kriterijumi za uključivanje studija u meta-analizu test-retest pouzdanosti u longitudinalnim studijama bili su da u studijama saopštene korelacije parova skorova na dva merenja, na istim ispitanicima u dva različita vremena za svaki par, da su ti skorovi dobijeni od strane istih ocenjivača i da je saopšten interval između parova merenja.

4.1.4.1. Analiza ocena test-retest pouzdanosti za HAMD skalu

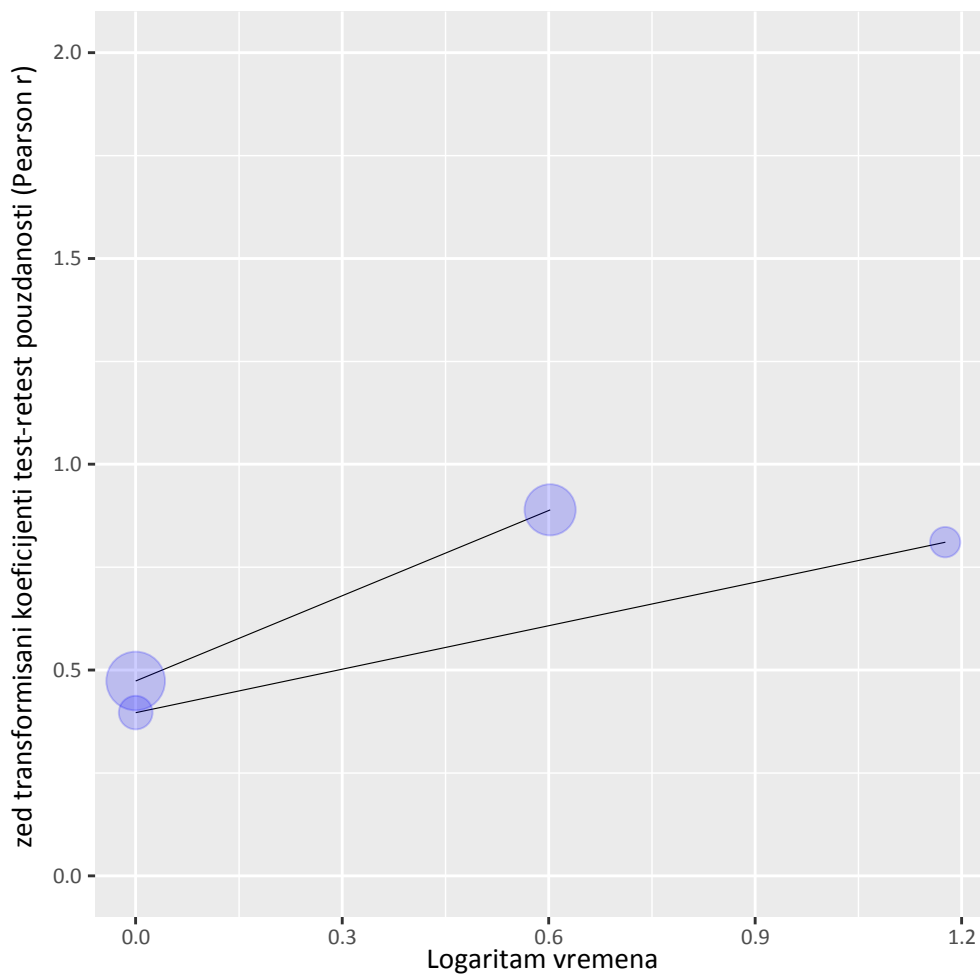
Ponovljena test-retest pouzdanost nije saopštena u nijednom radu sa skalama HAMD ili MADRS.

U dve studije (Kreuzer i sar, 2012; Pintor i sar 2006) saopšteni su IPD za HAMD skalu u bar 4 merenja pa su sekundarnom analizom izračunati Pearsonovi koeficijenti korelacije za prve dve, kao i za poslednje dve procene, što je test retest pouzdanost u longitudinalnim studijama (Prilog 5).

Ukupan broj ispitanika na prvoj proceni iznosio je 54. Obe studije bile su van engleskog govornog područja.

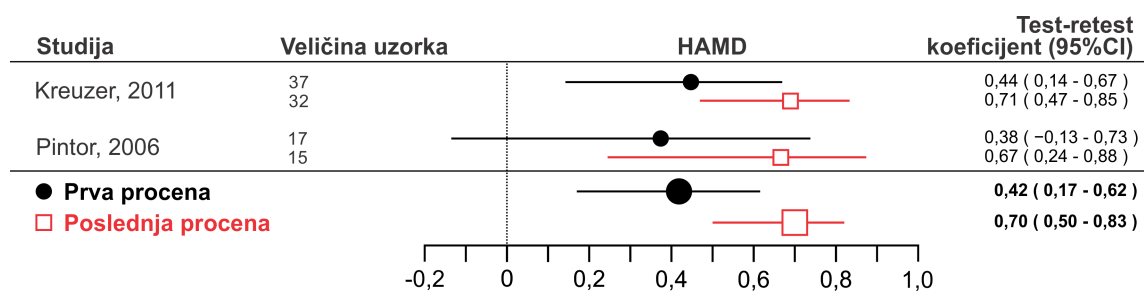
Postojanje kriterijuma za uključenje u studiju u vezi sa težinom depresije saopšteno je u studiji Pintor i sar (2006), pri čemu je korišćen HAMD skor viši od 8. Obe uključene studije po tipu su bile kliničke. U studiji Kreuzer i sar, 2012 korišćena je modifikovana verzija HAMD skale i ocenjivači su bili psihijatri. U studiji Pintor et al 2006 korišćena je originalna HAMD skala i ocenjivač bio psiholog.

Vrednost vremenskog intervala između parova korelacija iznosi od 1 do 15 dana.



Grafikon 22. Dijagram rasturanja sa individualnim linijama zeta transformisanih koeficijenata test retest pouzdanosti za HAMD skalu u odnosu na logaritam vremena procena depresivnosti unutar studija. Svaka puna linija predstavlja jednu studiju, a markeri predstavljaju zeta transformisane koeficijente test-retest pouzdanosti pojedinačnih procena tokom studije. Težinski koeficijent veličine markera je kvadratni koren iz recipročne varijanse zeta transformisanih koeficijenata test-retest pouzdanosti za HAMD skalu.

U poređenju sa početnim merenjima, vrednosti zed transformisanih koeficijenata test retest pouzdanosti za HAMD skalu više su na poslednjoj proceni kod obe studije uključene u meta-analizu. Objedinjena vrednost zed transformisanih koeficijenata test retest pouzdanosti za HAMD skalu iz modela slučajnih efekata su sledeće: 0,423 (95% CI 0,167 – 0,623) na prvom i 0,699 (95% CI 0,501 – 0,827) na poslednjem merenju.



Grafikon 23. Forest plot sa vrednostima koeficijenta test-retest pouzdanosti za HAMD skalu i pripadajućim 95% intervalima poverenja iz pojedinačnih studija na prvom i poslednjoj proceni.

4.2. Objedinjavanje valjanosti u vremenu

U cilju dobijanja objedinjenih ocena pouzdanosti urađena je analiza: koeficijenata korelacije, dijagnostičke i faktorske valjanosti za skale HAMD i MADRS.

4.2.1. Analiza ocena korelacija

Identifikovana su 62 rada koji ispunjavaju ove kriterijume za analizu korelacija: 31 rad je saopštio ponovljene koeficijente korelacije i 31 rad je saopštio individualne skore pacijenata (Prilog 6). Jedna studija sa IPD (Schlaepfer i sar, 2013) je isključena jer je koristila verzije HAMD skale sa različitim brojem ajtema i u različitim prilikama. Koeficijent korelacije je ekstrahovan iz studija ili je izračunat iz IPD. Lista studija uključenih u tri meta-analize korelacija, njihove karakteristike i koeficijenti korelacije između skala HAMD i BDI, HAMD i MADRS kao i između MADRS i BDI prikazana je u Prilogu 6. Od radova u kojima je saopšten koeficijent korelacije, Spearmanov koeficijent korelacije upotrebljen je u dve studije (Bowen, 1978; Flint i Rifat, 1996), dok je Pearsonov koeficijent saopšten u svim ostalim studijama. Za izračunavanje korelacija između skala iz IPD korišćen je Pearsonov koeficijent linearne korelacije.

Koeficijent korelacije između tri skale za procenu depresije ekstrahovan je ili sekundarnom analizom izračunat iz IPD iz 61 longitudinalne studije (1,6% od ukupno 3723 prikupljenih studija). Koeficijent korelacije između skala HAMD i BDI dobijen je iz 44 rada, korelacioni koeficijent između skala HAMD i MADRS dobijen je iz 18 radova, dok je koeficijent korelacije između skala MADRS i BDI dobijen iz 9 radova. Zbirne karakteristike studija uključenih u tri meta-analize korelacija skala HAMD i BDI, HAMD i MADRS, kao i MADRS i BDI prikazane su u Tabeli 12.

Veličina uzorka na početku studija je u opsegu 4 do 811. Medijana broja ispitanika najmanja je u studijama sa korelacijama HAMD i BDI skala. Veličina uzorka na kraju studija manja je nego na početku, što je i očekivano za longitudinalne studije. Ukupan broj ispitanika smanjuje se u sve tri meta-analize: za 9% u analizi skala HAMD i BDI, za 17% u analizi skala HAMD i MADRS i za 2% u analizi skala MADRS i BDI.

Trajanje studija kretalo se od 14 do 420 dana. Studije u kojima su obrađivane korelacije HAMD i BDI skala (medijana iznosi 30 dana) imaju upola kraće trajanje od studija u kojima su obrađivane ostale dve korelacije.

Depresija je najčešća primarna dijagnoza u svim studijama uključenim u meta-analizu. Postojanje kriterijuma za uključanje u studiju koji je u vezi sa težinom bolesti najčešće je korišćen (61%) kod studija sa korelacijama skala HAMD i MADRS. Prosečna starost ispitanika po studijama kretala se od 16,5 do 73,9 godina.

Objedinjene aritmetičke sredine psihijatrijskih skala su u skoro svim merenjima za 50% manje na kraju nego na početku studija. Najveće smanjenje objedinjenih aritmetičkih sredina zabeleženo je na MADRS skali, i to u obe korelacije. Varijabilitet je bio nešto veći na poslednjim merenjima. Najveći varijabilitet, u obe korelacije, na prvom i poslednjem merenju zabeležen je kod BDI skale.

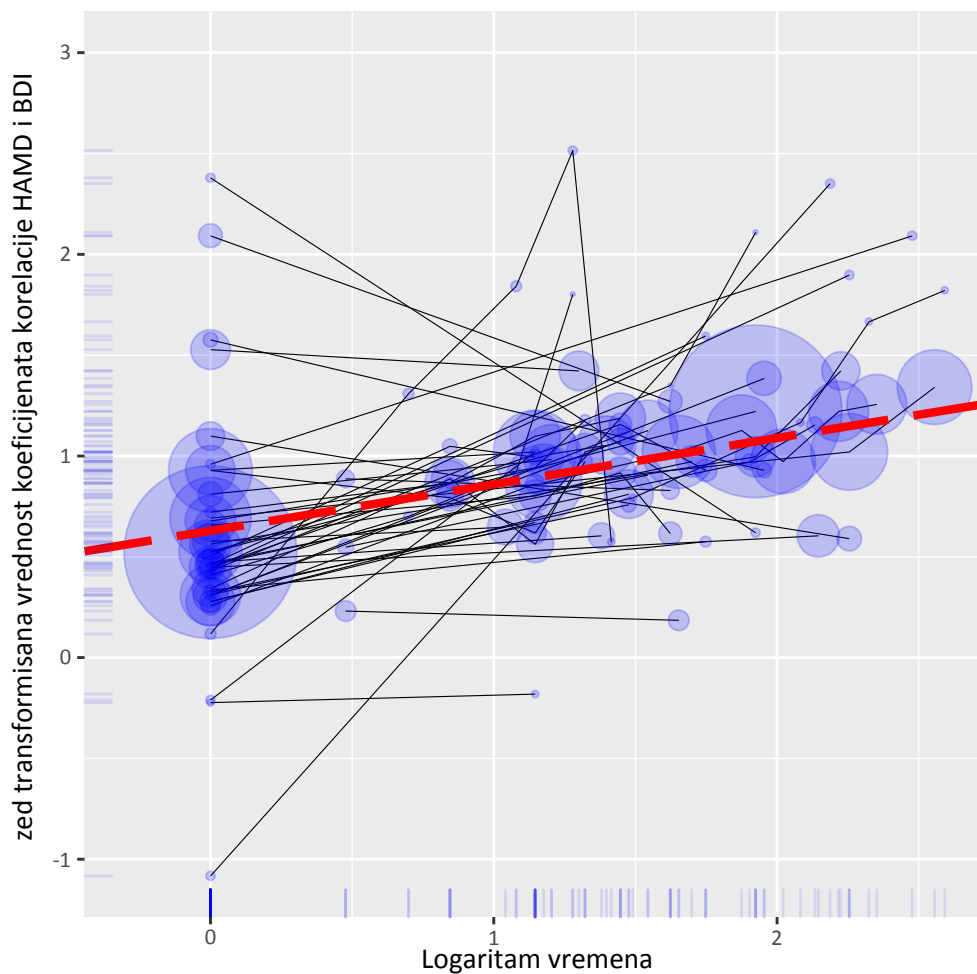
Pojedinačni zed-transformisani koeficijenti korelacije između skala HAMD i BDI, između HAMD i MADRS i između MADRS i BDI u svim merenjima prikazani su na Grafikonima 24, 25 i 26. Na ordinati su prikazani koeficijenti korelacije kao zed transformisane vrednosti, a na apscisi logaritama vremena procena depresivnosti unutar studija. Na ovim grafikonima svaka puna linija predstavlja jednu studiju, a markeri predstavljaju koeficijente korelacije pojedinačnih procena tokom studije. Objedinjena regresiona linija prikazana je zadebljanom isprekidanom linijom.

Broj procena depresivnosti primenom skala HAMD, MADRS ili BDI unutar studija kreće se od 2 do 13. Uzeto za sve studije zajedno, ispitanici su najčešće ocenjivani na početku i na kraju studije (63%). U studijama sa više od dva merenja mogu se zapaziti i oscilacije u vrednostima ispitivanih korelacija, što se lakše uočava u studijama sa većim brojem procena i dužim trajanjem. Objedinjena regresiona linija svih studija uključenih u pojedinačnu meta-analizu dobijena je iz regresionih koeficijenata koristeći RVE metod. Na Grafikonima 24, 25 i 26 vidi se da korelacije između skala rastu tokom vremena. Statistički značajne promene u vremenu nađene su za koeficijente korelacije između skala HAMD i BDI ($b=0,230$; $p<0,001$) i za koeficijente korelacije između skala HAMD i MADRS ($b=0,418$; $p<0,001$). Promene koeficijenata korelacije između skala MADRS i BDI su blizu statističke značajnosti ($b=0,166$; $p=0,069$), najverovatnije zbog malog uzorka.

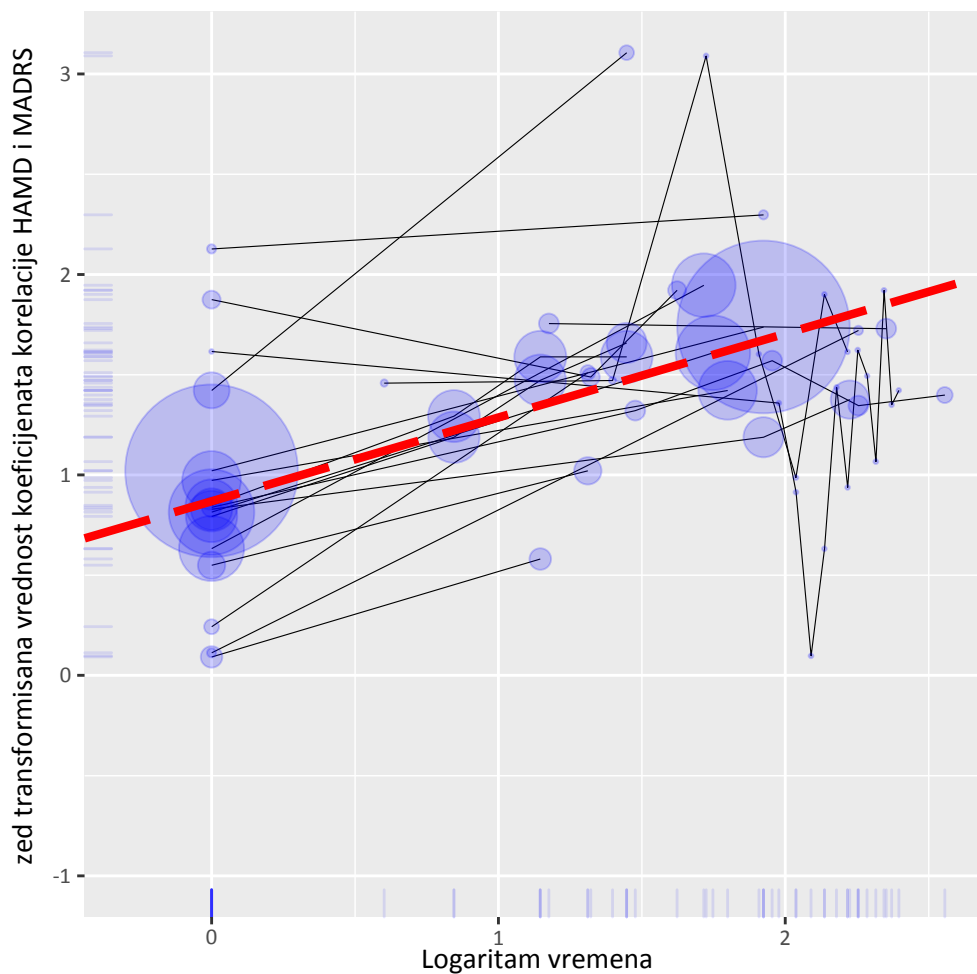
Tabela 12. Zbirne karakteristike studija uključenih u analizu korelacija skala HAMD, MADRS i BDI

Karakteristike studija	Koeficijenti korelacije između skala					
	HAMD i BDI		HAMD i MADRS		MADRS i BDI	
	Prva procena	Poslednja procena	Prva procena	Poslednja procena	Prva procena	Poslednja procena
Broj studija	44		18		9	
Ukupan broj ispitanika	2445	2227	1600	1330	1152	1125
Veličina uzorka, medijana (IQR)	19 (7-49)	18 (7-41)	21 (11-79)	16 (10-72)	17 (10-96)	17 (10-92)
Period praćenja (dani), medijana (IQR)		42 (21-90)		56 (28-168)		84 (28-120)
Depresija kao primarna dijagnoza, n (%)	34 (77)		17 (94)		7 (78)	
Kriterijum za uključenje u studiju u vezi sa težinom depresije, n (%)	18 (41)		11 (61)		3 (33)	
Starost (godine), AS (SD)	40,6 (10,9)		50,8 (12,0)		48,7 (6,0)	
HAMD skorovi						
Objedinjena AS	21,2	11,2	23,9	11,4		
Objedinjena SD	5,7	8,2	5,2	7,1		
BDI skorovi						
Objedinjena AS	24,2	11,4			25,8	13,1
Objedinjena SD	9,4	10,3			9,5	11,1
MADRS skorovi						
Objedinjena AS			29,6	12,0	25,3	11,1
Objedinjena SD			6,6	9,3	6,7	8,1
Poreklo koeficijenata korelacije, n (%)						
Publikovani u studiji	22 (50)		9 (50)		4 (44)	
Sekundarna analiza iz IPD	22 (50)		9 (50)		5 (56)	

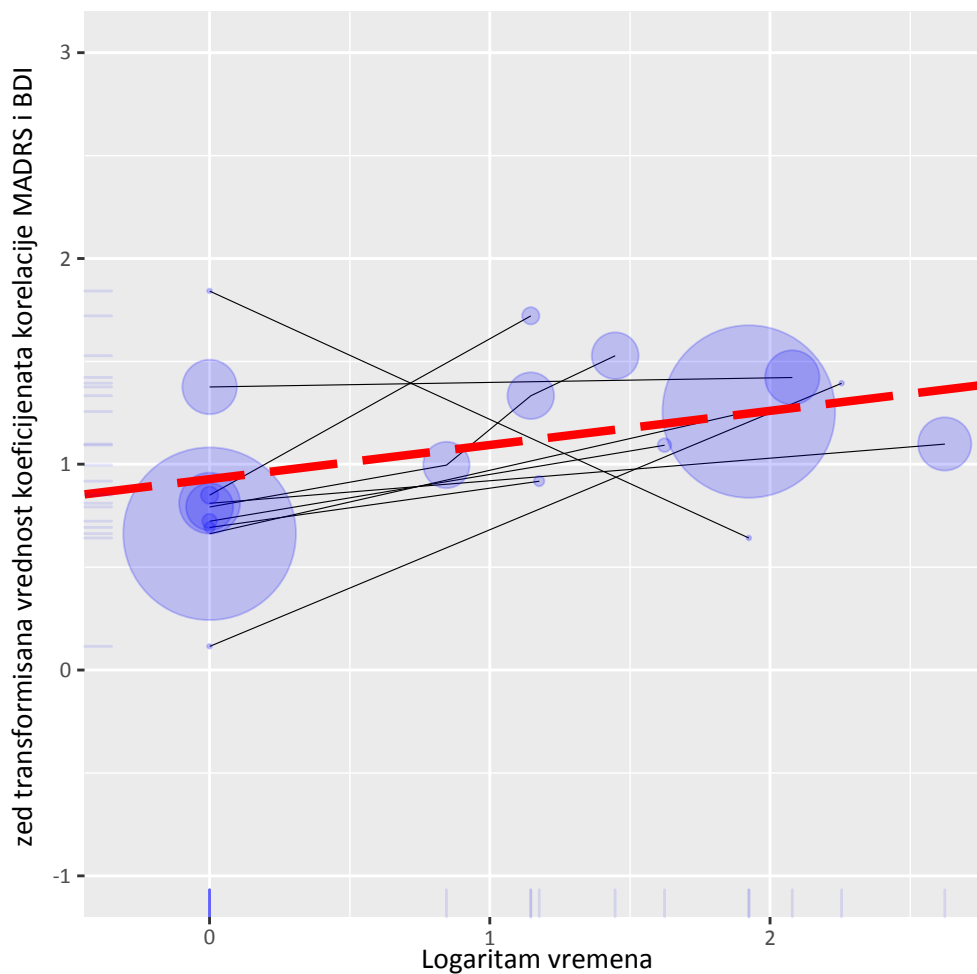
HAMD: Hamiltonova skala za procenu depresivnosti; MADRS: Montgomery-Ašbergova skala za procenu depresivnosti; BDI: Bekova skala za procenu depresivnosti; IQR: Interkvartilni opseg; IPD: Individualni podaci pacijenta; AS: Aritmetička sredina; SD: Standardna devijacija.



Grafikon 24. Dijagram rasturanja sa individualnim linijama zed transformisanih koeficijenata korelacije između skala HAMD i BDI u odnosu na logaritam vremena procena depresivnosti unutar studija. Svaka puna linija predstavlja jednu studiju, a markeri predstavljaju zed transformisane koeficijente korelacije pojedinačnih procena tokom studije. Težinski koeficijent veličine markera je kvadratni koren iz recipročne varijanse zed transformisanih koeficijenata korelacije. Objedinjena regresiona linija prikazana je zadebljanom isprekidanom linijom.



Grafikon 25. Dijagram rasturanja sa individualnim linijama zēd transformisanih koeficijenata korelacije između skala HAMD i MADRS u odnosu na logaritam vremena procena depresivnosti unutar studija. Svaka puna linija predstavlja jednu studiju, a markeri predstavljaju zēd transformisane koeficijente korelacije pojedinačnih procena tokom studije. Težinski koeficijent veličine markera je kvadratni koren iz recipročne varijanse zēd transformisanih koeficijenata korelacije. Objedinjena regresiona linija prikazana je zadebljanom isprekidanom linijom.



Grafikon 26. Dijagram rasturanja sa individualnim linijama zed transformisanih koeficijenata korelacije između skala MADRS i BDI u odnosu na logaritam vremena procena depresivnosti unutar studija. Svaka puna linija predstavlja jednu studiju, a markeri predstavljaju zed transformisane koeficijente korelacije pojedinačnih procena tokom studije. Težinski koeficijent veličine markera je kvadratni koren iz recipročne varijanse zed transformisanih koeficijenata korelacije. Objedinjena regresiona linija prikazana je zadebljanom isprekidanom linijom.

Analiza homogenosti pokazala je visok stepen heterogenosti sve tri ispitivane korelacije između skala: HAMD i BDI ($\tau^2=0,128$; $I^2=84\%$), HAMD i MADRS ($\tau^2=0,259$; $I^2=94\%$) i za MADRS i BDI ($\tau^2=0,193$; $I^2=92\%$), pa se pristupilo analizi moderatora primenom meta-regresije.

Meta-regersiona analiza izvedena je sa zed transformisanim koeficijentima korelacije kao zavisnim varijablama. U univarijantnim meta-regresionim modelima, aritmetičke sredine HAMD i BDI skorova su negativno povezane sa koeficijentima korelacije između skala HAMD i BDI. Suprotno od toga, vreme merenja i standardne devijacije skorova HAMD i BDI su pozitivno povezane sa koeficijentima korelacije između ovih skala. Shodno tome, kasnije ponovljene procene depresije i viši varijabilitet skala HAMD i BDI skorova povezan je sa višim koeficijentima korelacije. U multivarijantnom meta-regresionom modelu, standardna devijacija HAMD ukupnog skora je jedini statistički značajan prediktor zed transformisanih koeficijenata korelacije između skala HAMD i BDI ($p=0,040$) (Tabela 13).

Tabela 13. Univarijantni i multivarijantni meta-regresioni modeli za zed transformisanim koeficijentima korelacije između skala HAMD i BDI kao zavisnom varijablom

Nezavisna varijabla	Univarijantna meta-regresija		Multivarijantna meta-regresija	
	b	p	b	p
Primarna dijagnoza ispitanika (depresija ili drugi poremećaji)	0,208	0,245		
Kriterijum za uključenje u studiju u vezi sa težinom depresije	0,065	0.482		
Starost ispitanika (godine)	-0,009	0.219		
Logaritam vremena procena depresivnosti	0,234	<0,001		
HAMD prosečan skor	-0,029	0,002	-0,004	0,847
HAMD standardna devijacija	0,145	<0,001	0,088	0,040
BDI prosečan skor	-0,032	0,007	-0,017	0,427
BDI standardna devijacija	0,132	0,001	0,054	0,159

Tabela 14. Univarijantni meta-regresioni modeli za zed transformisanim koeficijentima korelacije između skala HAMD i MADRS kao zavisnom varijablom.

Nezavisna varijabla	Univarijantna meta-regresija	
	b	p
Kriterijum za uključenje u studiju u vezi sa težinom depresije	-0,077	0,776
Starost ispitanika (godine)	0,003	0,830
Logaritam vremena procena depresivnosti	0,428	<0,001
HAMD prosečan skor	-0,043	<0,001
HAMD standardna devijacija	0,162	0,025
MADRS prosečan skor	-0,017	0,028
MADRS standardna devijacija	-0,006	0,604

Druga meta-regresiona analiza pokazala je negativnu povezanost između aritmetičkih sredina HAMD i MADRS ukupnih skorova i koeficijenta korelacije između ovih skala. Drugim rečima, niže aritmetičke sredine skorova na obe skale su povezane sa višim koeficijentima korelacije između skala HAMD i MADRS. Suprotno od toga, vreme procene i standardna devijacija skorova na HAMD skali pozitivno je povezana sa koeficijentima korelacije između skala HAMD i MADRS. Varijabilitet skorova na MADRS skali nije povezan sa koeficijentima korelacije između ove dve skale.

Na Grafikonima 27, 28 i 29 prikazani su forest plotovi koeficijenata korelacije između skala HAMD i BDI, između HAMD i MADRS i između MADRS i BDI, sa pripadajućim 95% intervalima poverenja iz pojedinačnih studija, na prvoj i poslednjoj proceni unutar svake studije, respektabilno. U poređenju sa početnim merenjima, koeficijenti korelacije između skala za procenu depresije su više na poslednjim merenjima kod 52 (85%) studija uključenih u sve tri meta-analize korelacija.

Objedinjena vrednost koeficijenta korelacije iz modela slučajnih efekata je viša na poslednjem merenju u sve tri meta-analize. Najjača korelacija je nađena između skala

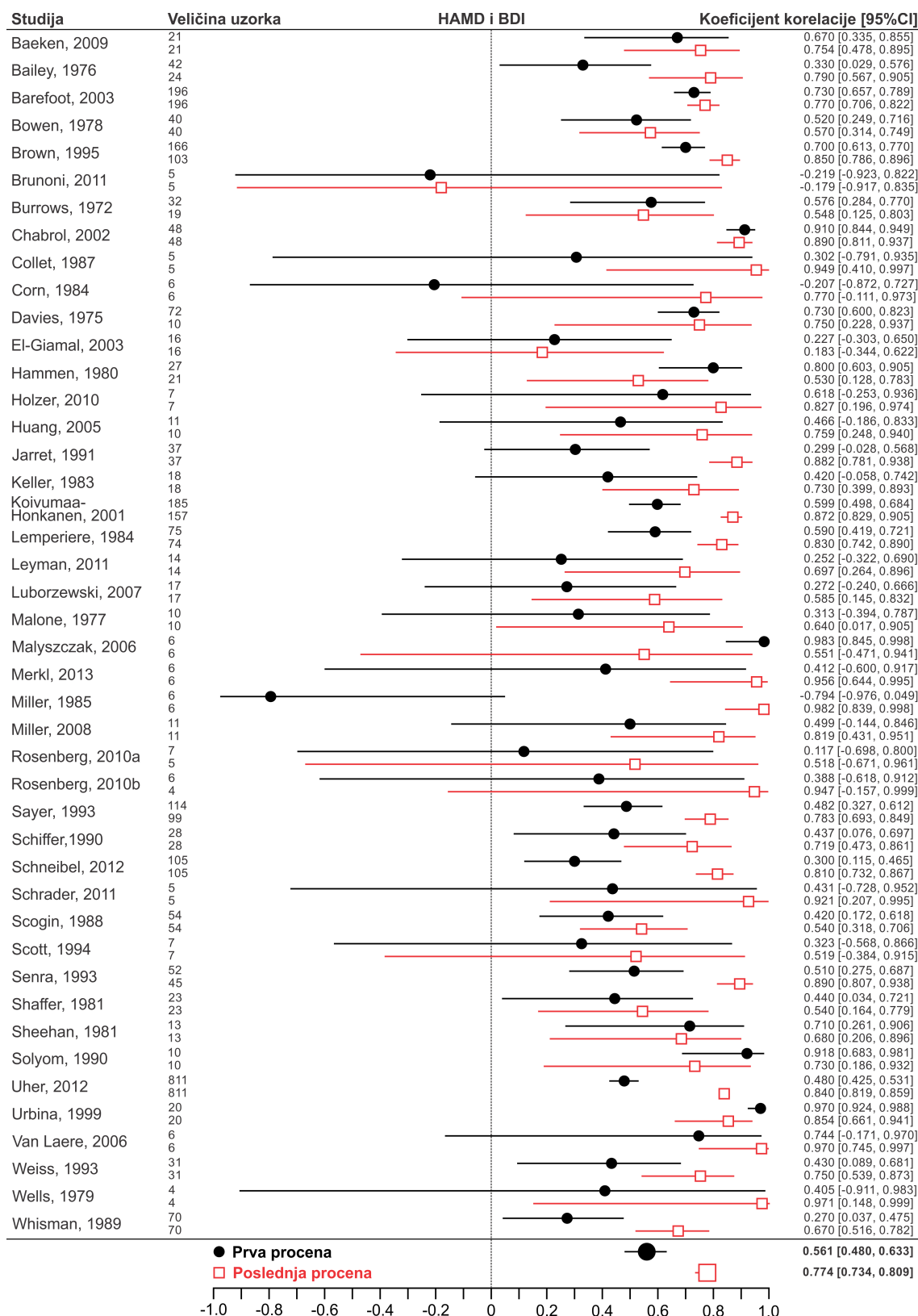
HAMD i MADRS, kako na prvom 0,722 (95% CI 0,645 - 0,785) tako i na poslednjem merenju 0,922 (95% CI 0,895 - 0,942). Objedinjene vrednosti koeficijenta korelacije između skala MADRS i BDI su sledeće: 0,704 (95% CI 0,564 - 0,804) na prvom i 0,863 (95% CI 0,825 - 0,894) na poslednjem merenju. Najniže objedinjene vrednosti koeficijenta korelacije su između skala HAMD i BDI, kako na početku 0,561 (95% CI 0,480 - 0,633) tako i na poslednjem merenju 0,774 (95% CI 0,734 - 0,809).

Publikaciona pristrasnost nije nađena niti primenom levkastog dijagrama niti primenom Eggerovog testa. Levkasti dijagrami na prvom i poslednjem merenju izgledaju dovoljno simetrično (Grafikoni 27, 28 i 29), a Eggerov test nije bio statistički značajan ($p > 0.05$) za sve tri ispitivane korelacije.

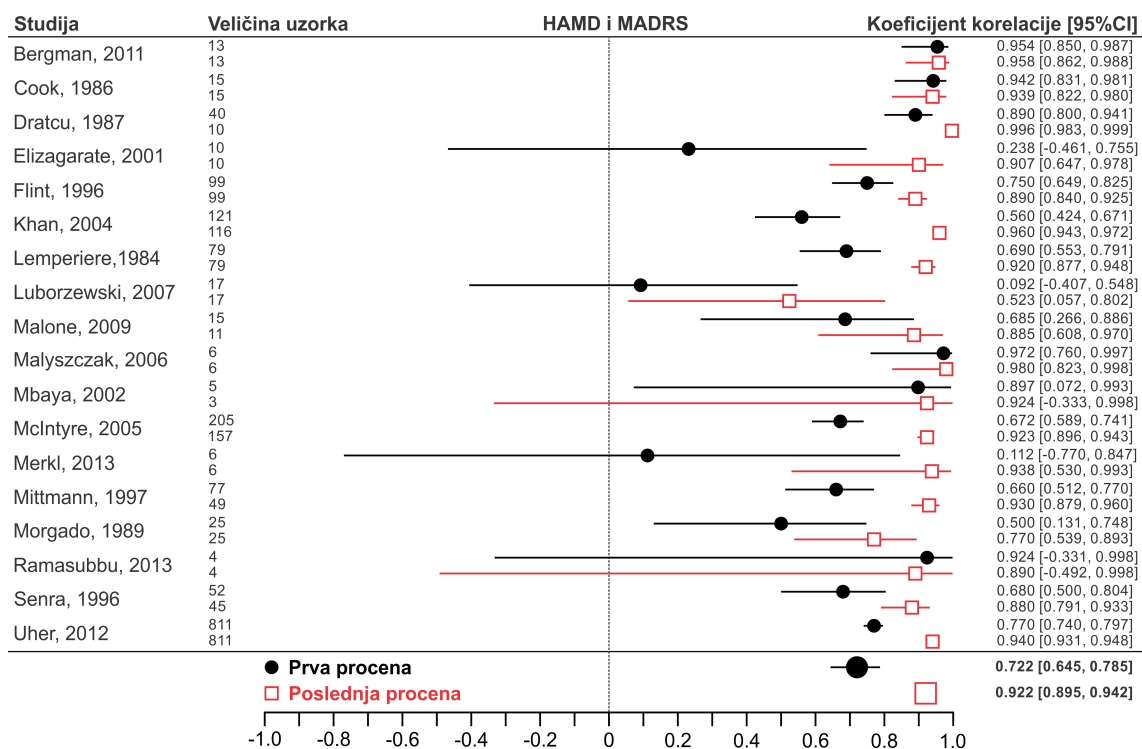
Na prvoj proceni zed transformisanih koeficijenta korelacije skala HAMD i BDI, Q test heterogenosti bio je statistički značajan ($p < 0,001$), $\tau^2 = 0,074$ i $I^2 = 77\%$, što ukazuje na visok stepen heterogenosti. Na poslednjoj proceni Q test heterogenosti bio je statistički značajan ($p < 0,001$), $\tau^2 = 0,040$ i $I^2 = 61\%$ što ukazuje na srednji stepen heterogenosti. Izgled Baujatovih dijagrama na prvoj i poslednjoj proceni, takođe ukazuje na heterogenost studija.

Na prvoj proceni zed transformisanih koeficijenta korelacije skala HAMD i MADRS, Q test heterogenosti bio je statistički značajan ($p < 0,001$), $\tau^2 = 0,051$ i $I^2 = 76\%$, što ukazuje na visok stepen heterogenosti. Na poslednjoj proceni Q test heterogenosti bio je statistički značajan ($p < 0,001$), $\tau^2 = 0,053$ i $I^2 = 74\%$ što ukazuje na srednji stepen heterogenosti. Izgled Baujatovih dijagrama na prvoj i poslednjoj proceni, takođe ukazuje na heterogenost studija.

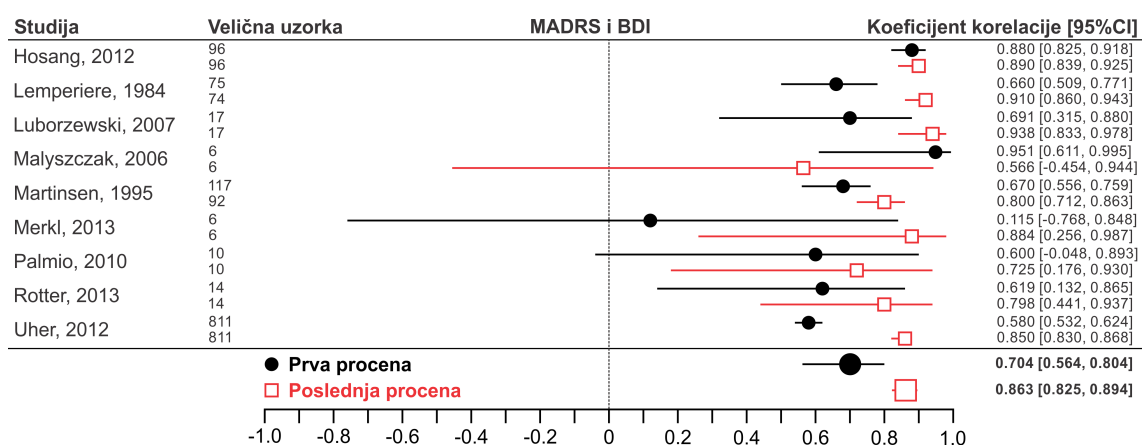
Na prvoj proceni zed transformisanih koeficijenta korelacije skala MADRS i BDI, Q test heterogenosti bio je statistički značajan ($p < 0,001$), $\tau^2 = 0,077$ i $I^2 = 83\%$, što ukazuje na visok stepen heterogenosti. Na poslednjoj proceni Q test heterogenosti nije bio statistički značajan ($p = 0,062$), $\tau^2 = 0,014$ i $I^2 = 46\%$ što ukazuje na srednji stepen heterogenosti. Izgled Baujatovih dijagrama na prvoj i poslednjoj proceni, takođe ukazuje na heterogenost studija.



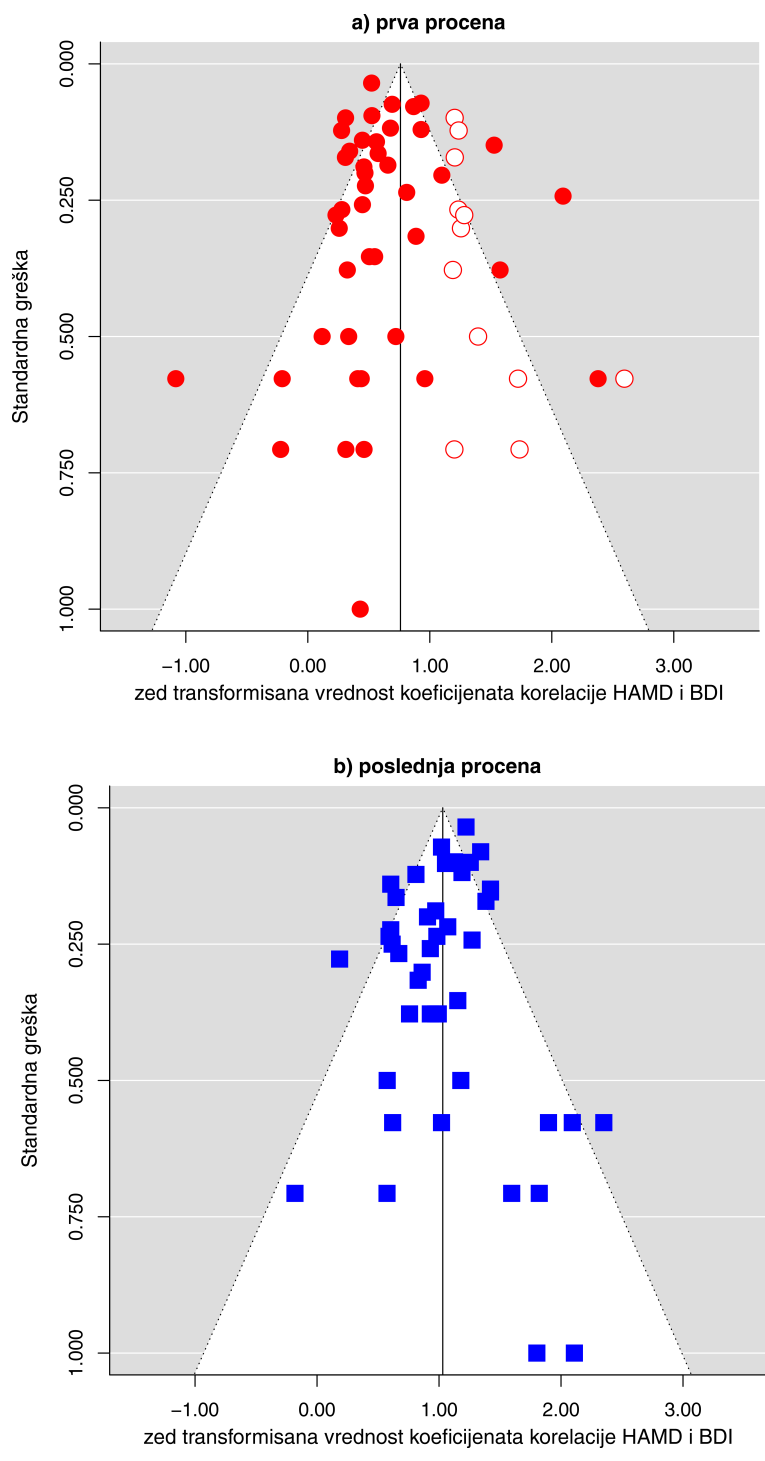
Grafikon 27. Forest plot sa vrednostima koeficijenta korelacije između skala HAMD i BDI sa pripadajućim 95% intervalima poverenja iz pojedinačnih studija na prvoj i poslednjoj proceni.



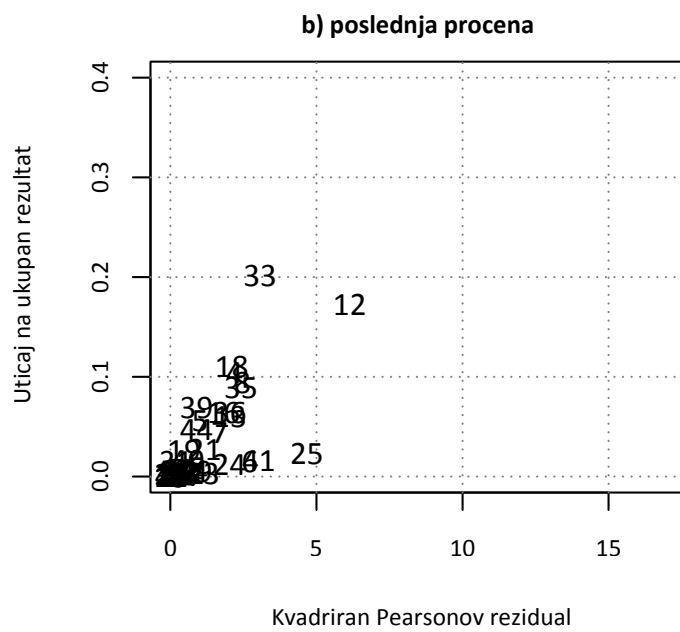
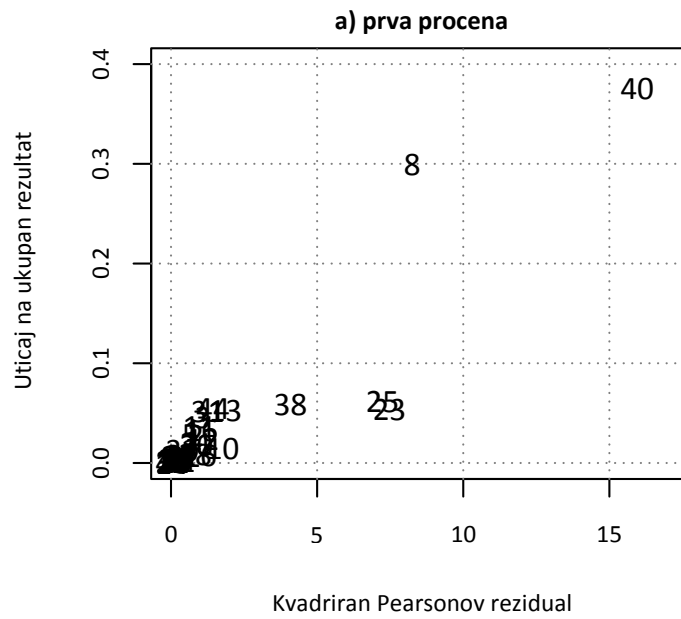
Grafikon 28. Forest plot sa vrednostima koeficijenta korelacije između skala HAMD i MADRS sa pripadajućim 95% intervalima poverenja iz pojedinačnih studija na prvoj i poslednjoj proceni.



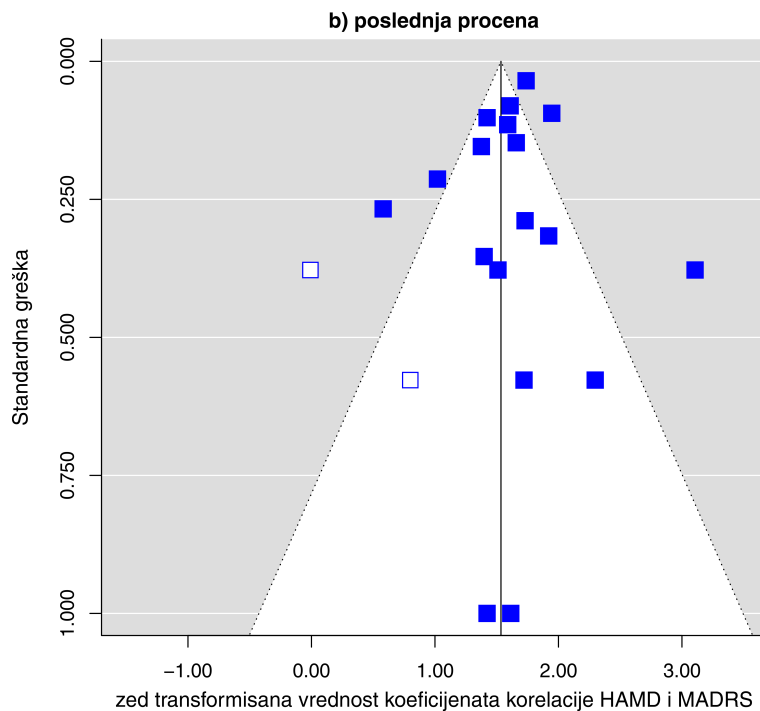
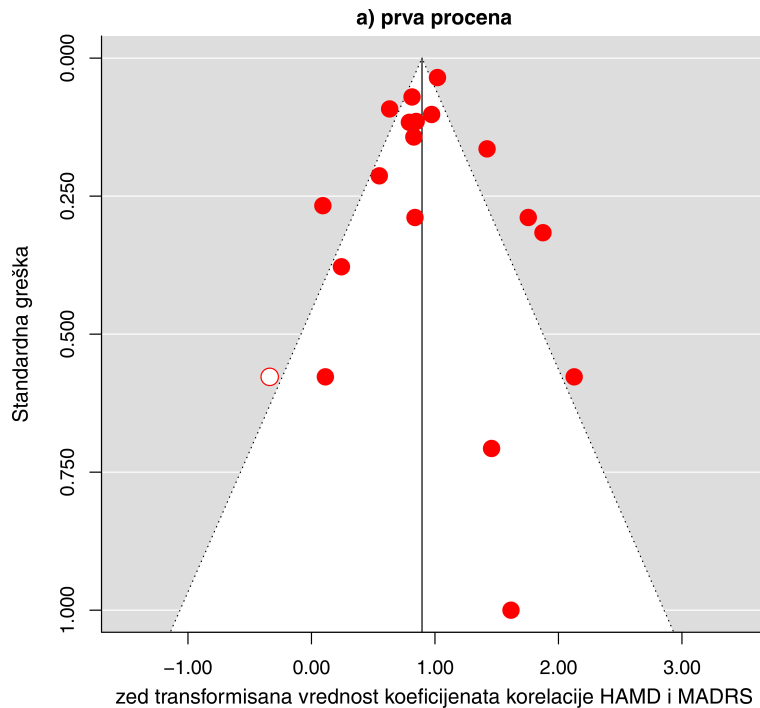
Grafikon 29. Forest plot sa vrednostima koeficijenta korelacije između skala MADRS i BDI sa pripadajućim 95% intervalima poverenja iz pojedinačnih studija na prvoj i poslednjoj proceni.



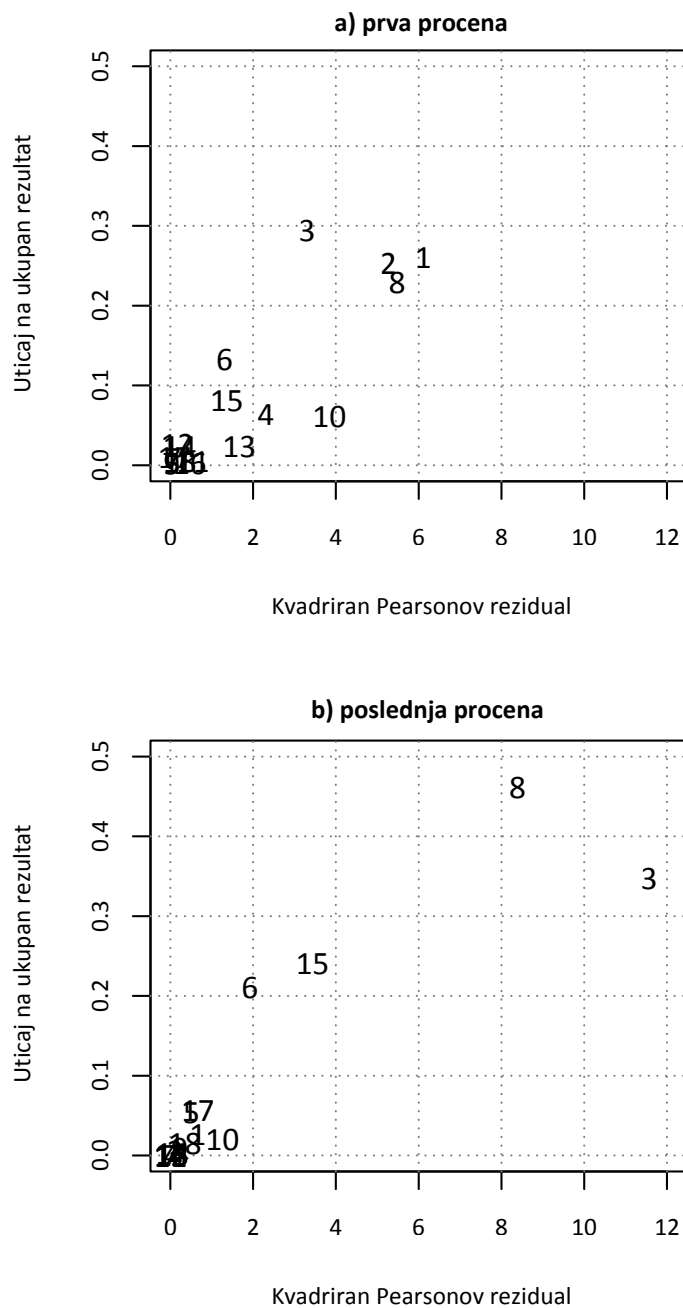
Grafikon 30. Levkasti dijagram (sa ukloniti i popuniti metodom dodavanja nedostajućih studija) zed transformisanih koeficijena korelacije između skala HAMD i BDI, na prvoj i poslednjoj proceni. Eggerov test: na prvoj proceni $p=0,309$; na poslednjoj proceni $p=0,863$.



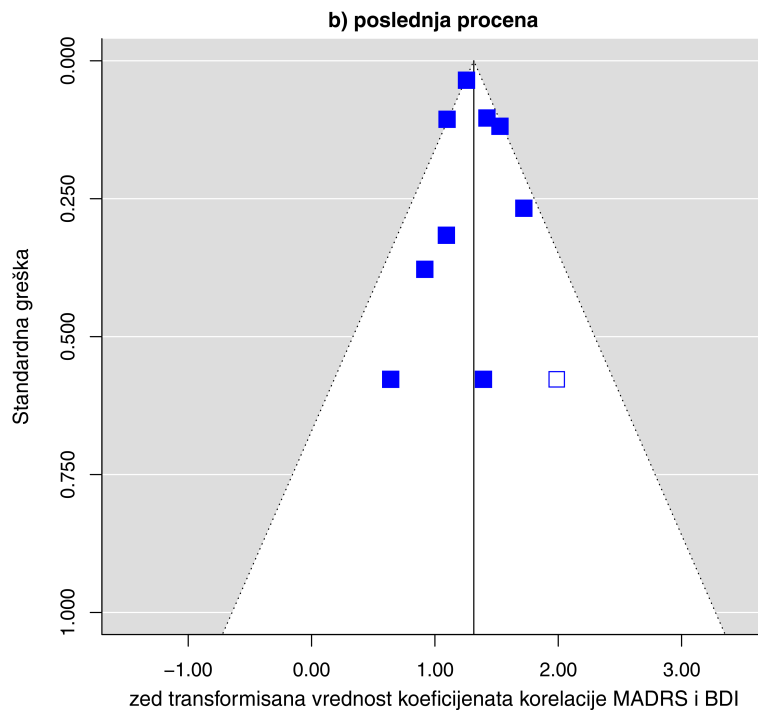
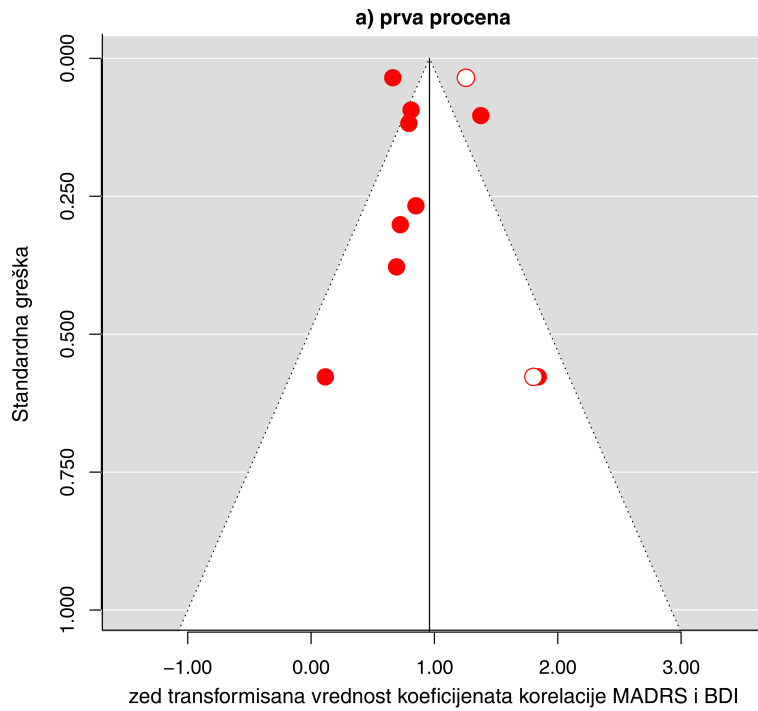
Grafikon 31. Baujatov dijagram za transformisanih koeficijenta korelacije skale HAMD i BDI na prvoj i poslednjoj proceni.



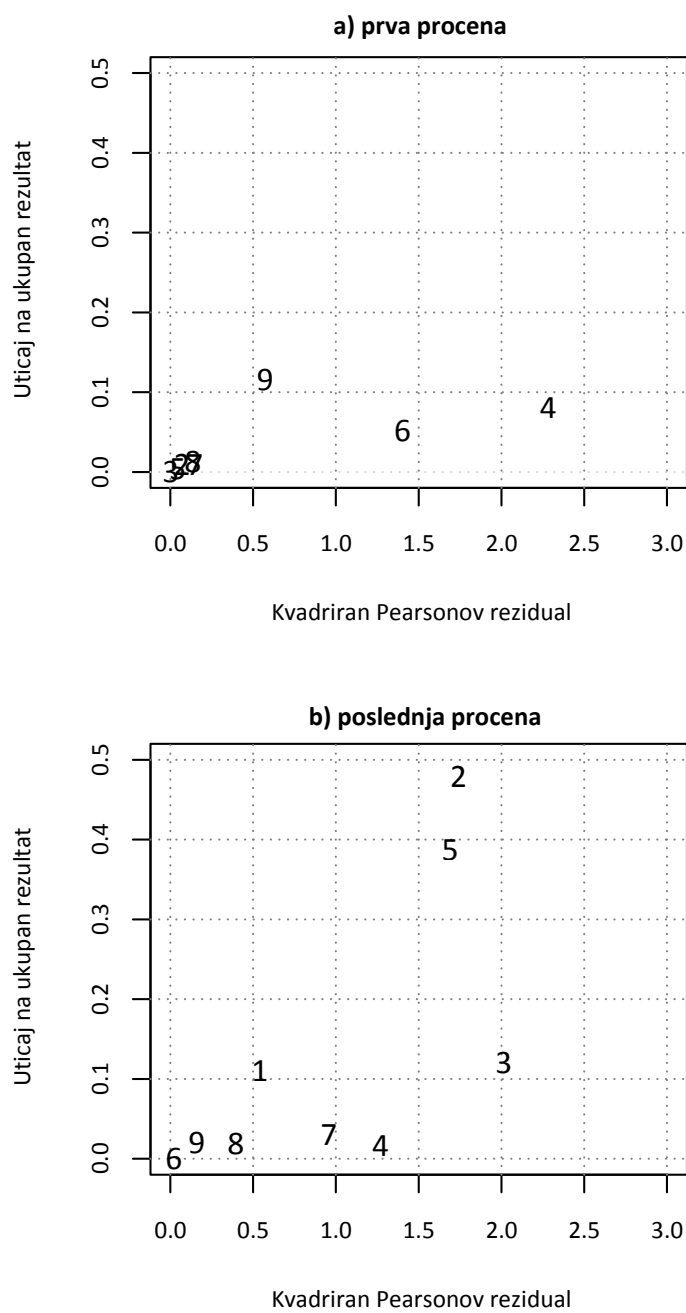
Grafikon 32. Levkasti dijagram (sa ukloniti i popuniti metodom dodavanja nedostajućih studija) zed transformisanih koeficijenata korelacije između skala HAMD i MADRS, na prvoj i poslednjoj proceni. Eggerov test: na prvoj proceni $p=0,303$; na poslednjoj proceni $p=0,642$.



Grafikon 33. Baujatov dijagram zed transformisanih koeficijenata korelacije skala HAMD i MADRS na prvoj i poslednjoj proceni.



Grafikon 34. Levkasti dijagram (sa ukloniti i popuniti metodom dodavanja nedostajućih studija) zed transformisanih koeficijena korelacije između skala MADRS i BDI, na prvoj i poslednjoj proceni. Eggerov test: na prvoj proceni $p=0,993$; na poslednjoj proceni $p=0,594$.



Grafikon 35. Baujatov dijagram zed transformisanih koeficijenata korelacije skala MADRS i BDI na prvoj i poslednjoj proceni.

4.2.2. Analiza dijagnostičke valjanosti

4.2.2.1. Analiza dijagnostičke valjanosti za HAMD skalu

Dijagnostička valjanost HAMD skale u ponovljenim merenjima identifikovana je u 4 studije (0,06% od ukupno 6590 prikupljenih studija).

Primarne dijagnoze kod uključenih studija bile su: moždani udar kod dve studije (Berg i sar, 2009; Kang i sar, 2013), pospartalni period (Thompson i sar, 1998) i zavisnost od kokaina (Weiss i sar, 1989). Primarnu dijagnozu depresije nije imala nijedna od uključenih studija.

Kriterijumski standard kod 3 studije bio je DSM i to DSM-III, DSM-III-R ili DSM-IV pojedinačno. U jednoj studiji kriterijumski standard bio je *Istraživački dijagnostički kriterijum* (RDC).

Iz studija uključenih u meta-analizu ekstrahovani su podaci o vremenu prve i poslednje procene, ukupnom broju ispitanika, primarnoj dijagnozi, statusu depresivnosti prema kriterijumskom standardu, dijagnostičkom pragu, kriterijumskom standardu i dijagnostičkoj valjanosti HAMD skale. Karakteristike studija uključenih u meta-analizu i rekonstruisane učestalosti dijagnostičke tabele prikazane su u Prilogu 7.

Od dijagnostičkih performansi HAMD skale ekstrahovani su podaci o: senzitivnosti, specifičnosti, pozitivnoj i negativnoj prediktivnoj vrednosti, površini ispod krive maksimalne verodostojnosti, odnosu verodostojnosti pozitivnog i negativnog testa. Jedan deo podataka je saopšten a jedan deo je izračunat sekundarnom analizom iz rekonstruisane dijagnostičke tabele. Dijagnostičke performanse uključenih studija prikazane su u prilogu 7.

Ukupan broj ispitanika na prvoj proceni iznosio je 770 a na poslednjoj 663. Medijana broja ispitanika na prvoj proceni iznosila je 128,5 (opseg, 90-423) dok je na poslednjoj proceni iznosila 142 (opseg, 91-288). Trajanje studija kretalo se u opsegu od 28 do 365 dana.

U dve studije je korišćena verzija HAMD skale sa 17 ajtema a u druge dve nije saopšten broj ajtema, međutim, pošto je citirana originalna referenca smatramo da je i to HAMD skale sa 17 ajtema.

U dve studije su saopštene više od jedne vrednosti dijagnostičkog praga, minimalne vrednosti HAMD skale za detektovanje depresije. Kod Berga i sar. (2009) za dijagnostičke pragove uzete su vrednosti HAMD skale više ili jednako 7, 10 i 12. U studiji Weiss i sar. (1989) vrednosti praga HAMD skale iznosile su više i jednako 7, 8, 10 i 17. U oba slučaja u analizu su uzete one vrednosti dijagnostičkog praga sa najvišom vrednošću zbira senzitivnosti i specifičnosti. U jednoj studiji je saopštena jedna vrednost a u jednoj nije uopšte saopštena vrednost dijagnostičkog praga. U svim studijama uključenim u meta-analizu upotrebljene su iste vrednosti dijagnostičkog praga na prvoj i poslednjoj proceni. Sve vrednosti dijagnostičkog praga su različite (10, 12 ili 17).

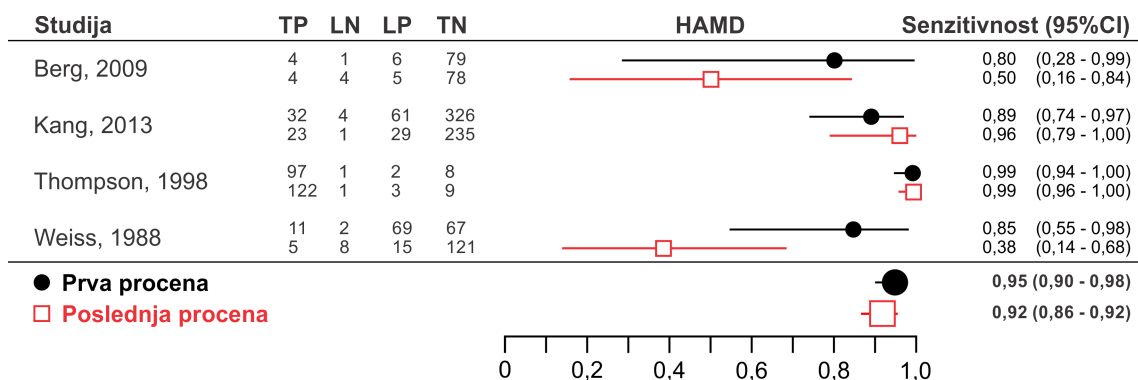
Prema statusu depresivnosti u odnosu na kriterijumski standard u tri studije poredile su major depresiju prema nedepresivnosti, dok je u jednoj studiji poređena depresija u odnosu na nedepresivnost.

U Tabeli 15 prikazane su objedinjene vrednosti ocena dijagnostičkih performansi na prvoj i poslednjoj proceni. Objedinjena vrednost senzitivnosti na prvoj proceni iznosila je 0,947 a na poslednjoj 0,917. Objedinjena vrednost specifičnosti na prvoj proceni iznosila je 0,777 a na poslednjoj 0,895. Objedinjena vrednost DOR iz modela slučajnih efekata na prvoj proceni iznosila je 38,391 a na poslednjoj 41,135. Za sve ocene na prvoj proceni izuzev LR- postojala je statistički značajna heterogenost između studija, dok kod poslednje proceni statistička značajnosti nije postojala kod specifičnosti i LR+.

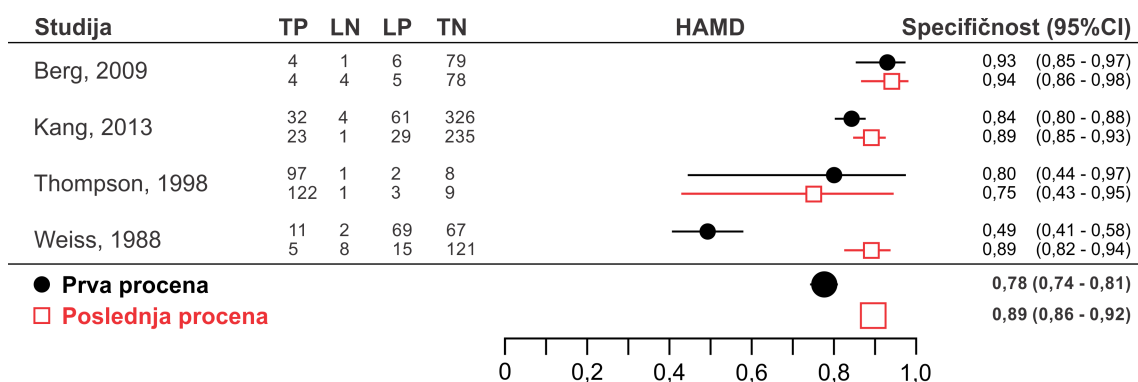
Tabela 15. Objedinjene ocene dijagnostičkih performansi HAMD skale iz svih studija na prvoj i poslednjoj proceni

Objedinjene ocene dijagnostičkih performansi	Prva procena	Poslednja procena
	Aritmetička sredina (95% IP)	Aritmetička sredina (95% IP)
Senzitivnost	0,947 (0,899 – 0,977) Hi-kvadrat=10,24; p=0,017	0,917 (0,864 – 0,954) Hi-kvadrat=48,03; p<0,001
Specifičnost	0,777 (0,742 – 0,809) Hi-kvadrat=77,26; p<0,001	0,895 (0,865 – 0,921) Hi-kvadrat=4,19; p=0,242
LR+	4,547 (1,844-11,211) Hi-kvadrat=47,71; p<0,001	5,922 (3,231-10,856) Hi-kvadrat=7,47; p=0,058
LR-	0,121 (0,040-0,367) Hi-kvadrat=7,42; p=0,060	0,153 (0,030-0,772) Hi-kvadrat=35,88; p<0,001
DOR	38,391 (7,909-186,35) Hi-kvadrat=9,74; p=0,021	41,135 (5,398-313,46) Hi-kvadrat=15,94; p=0,001

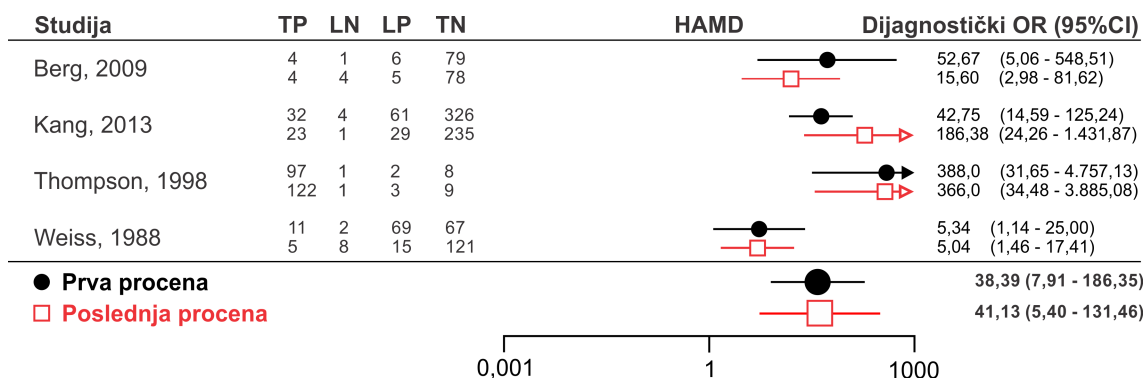
a) Senzitivnost



b) Specifičnost



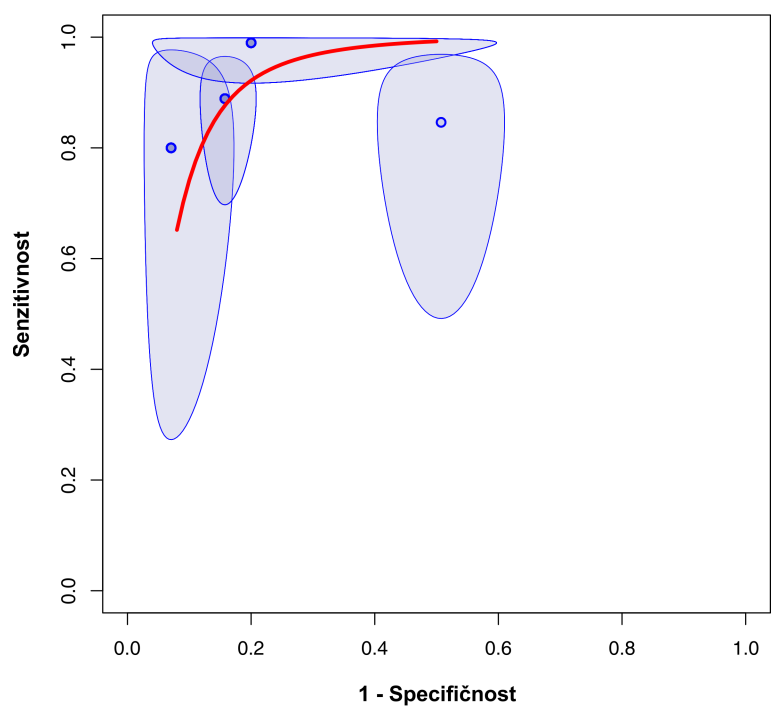
c) Dijagnostički odnos šansi



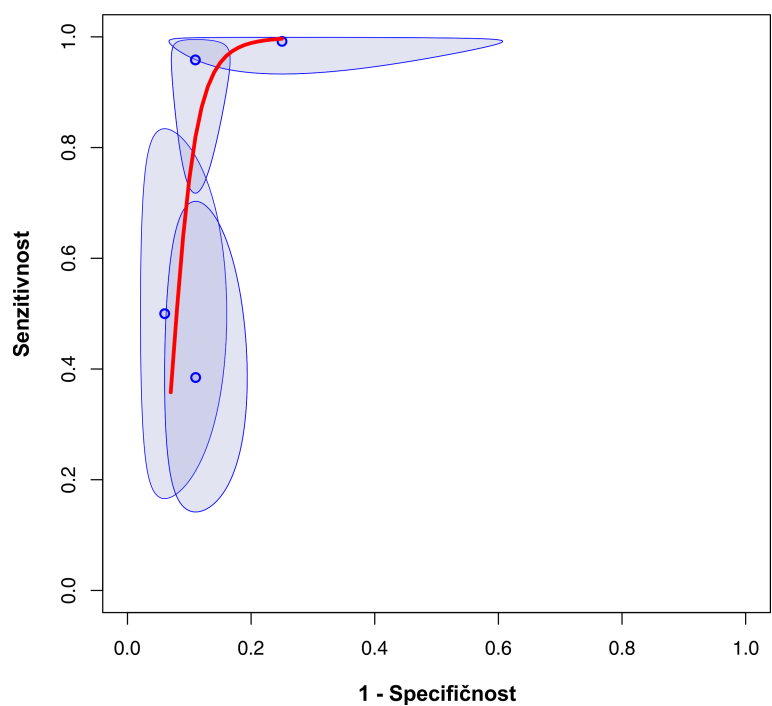
Grafikon 36. Forest plot ocena senzitivnosti, specifičnosti i dijagnostičkog odnosa šansi za HAMD skalu na prvoj i poslednjoj proceni.

Na prvoj proceni postojala je značajna korelacija $\rho=-0,013$ a na poslednjoj je iznosila $\rho=0,653$, što ukazuje na postojanje efekta dijagnostičkog praga na prvoj proceni.

a) prva procena



b) poslednja procena



Grafikon 37. Objedinjene krive maksimalne verodostojnosti (SROC) za HAMD skalu.

Zbog malog broja studija koje su uključene u meta-analizu, nije bilo moguće uraditi analizu senzitivnosti.

4.2.2.2. Analiza dijagnostičke valjanosti za MADRS skalu

Dijagnostička valjanost MADRS skale u ponovljenim merenjima identifikovana je u 2 studije (0,1% od ukupno 2051 prikupljenih studija).

U obe studije uključene u meta-analizu (Kang i sar, 2013; Laska i sar, 2007) primarna dijagnoza bila je moždani udar. Primarnu dijagnozu depresije nije imala nijedna od uključenih studija. Kriterijumski standard u obe uključene studije bio je DSM-IV.

Iz studija uključenih u meta-analizu ekstrahovani su podaci o vremenu prve i poslednje procene, ukupnom broju ispitanika, primarnoj dijagnozi, statusu depresivnosti prema kriterijumskom standardu, dijagnostičkom pragu, kriterijumskom standardu i dijagnostičkoj valjanosti MADRS skale. Karakteristike studija uključenih u meta-analizu i rekonstruisane učestalosti dijagnostičke tabele prikazane su u Prilogu 7.

Od dijagnostičkih performansi MADRS skale ekstrahovani su podaci o: senzitivnosti, specifičnosti, pozitivnoj i negativnoj prediktivnoj vrednosti, površini ispod krive maksimalne verodostojnosti, odnosu verodostojnosti pozitivnog i negativnog testa. Jedan deo podataka je saopšten a drugi deo je izračunat sekundarnom analizom iz rekonstruisane dijagnostičke tabele. Dijagnostičke performanse uključenih studija prikazane su u Prilogu 7.

Ukupan broj ispitanika na prvoj proceni iznosio je 496 a na poslednjoj 365. Studije su trajale 180 dana (Laska i sar, 2007) i 365 dana (Kang i sar, 2013).

U obe studije je korišćena originalna verzija MADRS skale sa 10 ajtema.

U studijama su saopštene različite vrednosti dijagnostičkog praga, minimalne vrednosti MADRS skale za detektovanje depresije, više i jednako 8 ili 10. U svim studijama uključenim u meta-analizu upotrebljene su iste vrednosti dijagnostičkog praga na prvoj i poslednjoj proceni.

Prema statusu depresivnosti prema kriterijumskom standardu studija Kang i sar. (2013) poredila je major depresiju prema nedepresivnosti, dok u studiji Laska i sar. (2007) nisu navedeni podaci o statusu depresivnosti prema kriterijumskom standardu.

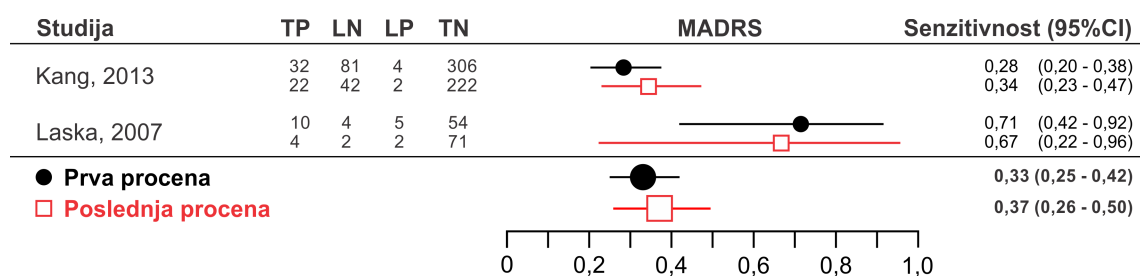
U Tabeli 16 prikazane su objedinjene vrednosti ocena dijagnostičkih performansi na prvoj i poslednjoj proceni. Objedinjena vrednost senzitivnosti na prvoj proceni iznosila

je 0,331 a na poslednjoj 0,371. Objedinjena vrednost specifičnosti na prvoj proceni iznosila je 0,976 a na poslednjoj 0,987. Objedinjena vrednost DOR iz modela slučajnih efekata na prvoj proceni iznosila je 29,076 a na poslednjoj 61,885. Za ocene senzitivnosti, specifičnosti i LR- na prvoj proceni postojala je statistički značajna heterogenost između studija, dok kod poslednje proceni statistička značajnost nije postojala kod nijedne ocene dijagnostičkih performansi MADRS skale.

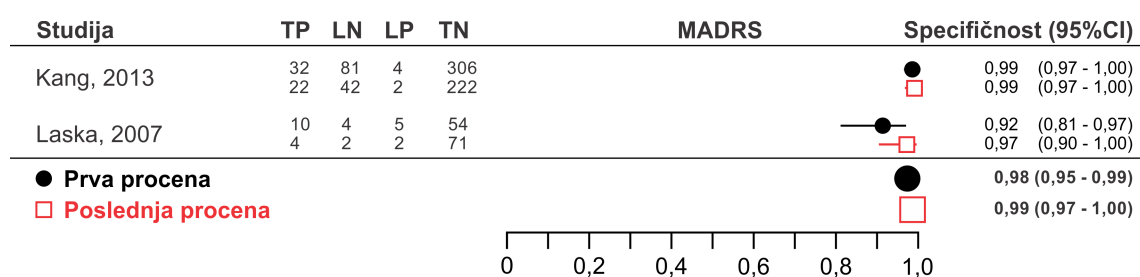
Tabela 16. Objedinjene ocene dijagnostičkih performansi MADRS skale iz svih studija na prvoj i poslednjoj proceni

Objedinjene ocene dijagnostičkih performansi	Prva procena	Poslednja procena
	Aritmetička sredina (95% IP)	Aritmetička sredina (95% IP)
Senzitivnost	0,331 (0,250 – 0,420) Hi-kvadrat=9,78; p=0,002	0,917 (0,864 – 0,954) Hi-kvadrat=48,03; p<0,001
Specifičnost	0,976 (0,954 – 0,989) Hi-kvadrat=7,63; p=0,006	0,895 (0,865 – 0,921) Hi-kvadrat=4,19; p=0,242
LR+	13,253 (4,829 – 36,376) Hi-kvadrat=2,21; p=0,137	5,922 (3,231-10,856) Hi-kvadrat=7,47; p=0,058
LR-	0,520 (0,215 – 0,717) Hi-kvadrat=4,60; p=0,032	0,153 (0,030-0,772) Hi-kvadrat=35,88; p<0,001
DOR	29,076 (12,236 – 69,094) Hi-kvadrat=0,02; p=0,902	41,135 (5,398-313,46) Hi-kvadrat=15,94; p=0,001

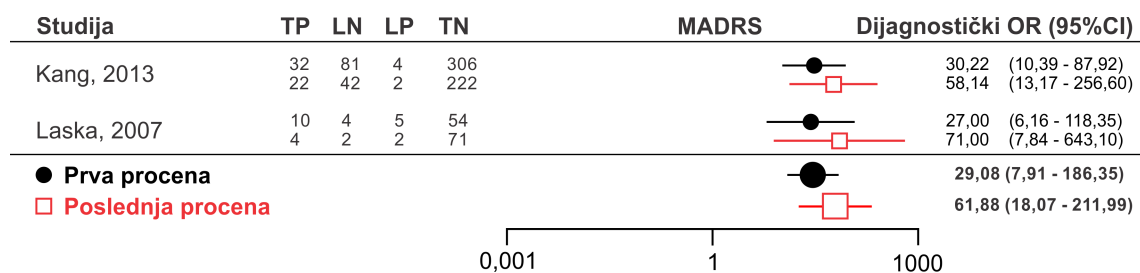
a) *Senzitivnost*



b) *Specifičnost*



c) *Dijagnostički odnos šansi*



Grafikon 38. Forest plot ocena senzitivnosti, specifičnosti i dijagnostičkog odnosa šansi za MADRS skalu na prvoj i poslednjoj proceni.

4.2.3. Analiza faktorske valjanosti

4.2.3.1. Analiza faktorske valjanosti za HAMD skalu

Faktorska valjanost HAMD skale u ponovljenim merenjima saopštena je u 4 studije (Prilog 8).

Tabela 17. Karakteristike studija uključenih u faktorsku analizu za HAMD skalu

Studija	vreme	n	Broj ekstra faktora	Ukupna objašnjena varijansa (%)	Naziv prvog faktora	Varijansa objašnjena prvim ekstrahovanim faktorom (%)
Addington, 1996	Time 1	112	7	65,5		21,3
	Time 2	89	7	71,8		26,5
Bent- Hansen, 2003	Baseline	230	6			12,4
	Week 3	204	6			22,9
Licht, 2005	Baseline	1629	7			14,6
	2 weeks	1581	6			17,5
	4 weeks	1459	5			27,8
Maier, 1985	P1	107	4	22,1	Inhibicija	
	P3	88	4	33,1		
Schennach, 2012	Admision	278	3	40		14
	Discharge	278	3	42		16
Steinmeyer, 1992	Baseline	223	6			
	Endpoint	174	2			

Najmanje povećanje varijanse na prvom faktoru zapaženo je na uzorku shizofrenih pacijenata (Schennach i sar, 2012).

Najveća razlika smanjenja broja faktora zapažena je na uzorku pacijenata od major depresije (Steinmeyer i Möller, 1992), gde je broj faktora sa šest na prvom merenju opao na dva drugom merenju.

Medijana broja ekstrahovanih faktora HAMD skale na prvom merenju iznosila je 6 (opseg, 4-7), dok je na poslednjem merenju medijana iznosila 5 (opseg, 2-7), što nije statistički značajna razlika ($V=3$; $p=0,371$).

Ukupna objašnjena varijansa faktora HAMD skale na prvom merenju iznosila je 40 % (opseg, 22,1-65,5), dok je na poslednjem merenju medijana iznosila 42% (opseg, 33,1-71,8), što nije statistički značajna razlika ($V=0$; $p=0,250$).

Varijansa objašnjena prvim faktorom HAMD skale na prvom merenju iznosila je 14,3 % (opseg, 12,4-21,3), dok je na poslednjem merenju medijana iznosila 24,7% (opseg, 16,0-27,8), što nije statistički značajna razlika ($V=0$; $p=0,125$). U testu predznaka $p=0,063$.

4.2.3.2. Analiza faktorske valjanosti za MADRS skalu

Faktorska valjanost MADRS skale u ponovljenim merenjima saopštena je u 4 studije (Prilog 8).

Tabela 18. Karakteristike studija uključenih u faktorsku analizu za MADRS skalu

Studija	vreme	n	Broj ekstra faktora	Ukupna objašnjena varijansa (%)	Naziv prvog faktora	Varijansa objašnjena prvim ekstrahovanim faktorom (%)
Bent-Hansen, 2003	Baseline	230	5			17,1
	Week 3	204	4			27,7
Bondolfi, 2010	Time 1	63	2			42,3
	(Study inclusion)	63	2			57,6
	Time 2 (4 weeks later)					
Galinowski,1995	Day 0	137	3		Opšti faktor	32,3 66,3
	Day 28	137	1			
Rocca, 2002	Baseline	118	2		Opšti faktor	32,9 37
	After 8 weeks	100	1			

Medijana broja ekstrahovanih faktora MADRS skale na prvom merenju iznosila je 2,5 (opseg, 1-5), dok je na poslednjem merenju medijana iznosila 1 (opseg, 1-4), što nije statistički značajna razlika ($V=6$; $p=0,174$). Test predznaka $p=0,087$.

Varijansa objašnjena prvim faktorom MADRS skale na prvom merenju iznosila je 32,6% (opseg, 17,1-42,3), dok je na poslednjem merenju medijana iznosila 47,3% (opseg, 27,7-66,3), što nije statistički značajna razlika ($V=0$; $p=0,125$). U testu predznaka $p=0,063$.

5. Diskusija

U meta-analizama interne konzistentnosti HAMD i MADRS skala, ispitivane su promene mera interne konzistentnosti tokom vremena pri ponovljenim merenjima. Glavni rezultat je da interna konzistentnost raste tokom vremena, što je pokazano objedinjavanjem ocena alfa-koeficijenta, Loevingerovog koeficijenta i ICC.

Interna konzistentnost se najčešće meri primenom Cronbachovog alfa koeficijenta. Vrednosti ovog koeficijenta veće od 0,7 uzimaju se kao dokaz unidimensionalnosti instrumenta, dok visoke vrednosti mogu ukazivati na redundantnost ajtema (Bech i sar., 1997; Kobak i sar., 2010). Drugi nedostatak ovog koeficijenta je što njegove vrednosti zavise od broja ajtema (Bech i sar., 1997). Ovo nameće potrebu primene Spearman-Brownove formule (Brown, 1910; Spearman, 1910) koja je i primenjena u ovom istraživanju za korigovanje alfa koeficijenta koji su se odnosili na verzije HAMD skala dužine različite od 17 ajtema.

U istraživanjima se uglavnom, za meru interne konzistentnosti, publikuju Cronbachovi alfa koeficijent iz pojedinačnih merenja. Od strane pojedinih autora (npr. van de Rest i sar., 2010) ukazano je da pojedinačna merenja mogu dati potcenjene vrednosti interne konzistentnosti instrumenata, zbog teškoće ispunjavanja svih potrebnih pretpostavki kao što je jednakost varijansi, i da je takve ocene adekvatnije dobiti na osnovu ponovljenih merenja.

Interna konzistentnost merena Cronbach alfa koeficijentom reflektuje prosečne korelacije između ajtema, i kao takvog može se očekivati njegovo povećanje sa povećanjem korelacija između ajtema (Kobak i sar., 2010; van de Rest i sar., 2010). U ovom istraživanju objedinjene vrednosti alfa koeficijenta su na početku studija, kako za HAMD (objedinjeno alfa 0,66) tako i za MADRS skalu (objedinjeno alfa 0,68), nešto ispod granične vrednosti od 0,7 (npr. Bech i sar., 1997; Stage i sar., 2003) koja se u literaturi najčešće navodi kao zadovoljavajući, odnosno prihvatljiv, nivo za psihijatrijske skale. Na kraju studija su objedinjene vrednosti Cronbachovih koeficijenta alfa na zadovoljavajućem nivou (objedinjeno alfa: HAMD 0,81, MADRS 0,86). Ovo ukazuje da

tokom vremena dolazi kako do povećanja korelacija između ajtema unutar skala, tako i do povećanja korelacija između skala.

Interpretacija objedinjenih vrednosti u izvesnoj meri zavisi od preporuka autora o njihovoj vrednosti. Prema jednom izvoru iz klasične psihijatrijske literature (Blacker, 2009) kriterijum kada se može smatrati da skala ima dobre vrednosti interne konzistenosti je alfa $>0,75$. Prema jednom od klasičnih izvora psihološke literature (Nunnally, 1978) prihvatljiva vrednosti alfa koeficijenata dovoljne za preliminarna istraživanja iznose $>0,7$, za bazična istraživanja alfa $>0,8$, i za primenjena istraživanja alfa $>0,9$. Prema najstrožim kriterijumima jednog od izvora iz klasične opšte medicinske literature (Bland i Altman, 1997) zadovoljavajuće vrednosti za potrebe komparacija grupa su alfa od $0,7-0,8$, a za kliničku primenu alfa najmanje $0,9$. Uzimajući u obzir ove preporuke, jasno je da su vrednosti objedinjenih alfa koeficijenta nešto niže na početku nego što je zahtevano, dok su na kraju studija uglavnom zadovoljavajuće. Kako se radi o skalama sa višedecenijskom upotrebom, i koje i pored pojave novih instrumenata i dalje predstavljaju najkorišćenije referentne instrumente za validaciju drugih instrumenata, nameće se potreba redefinisanja minimalno zahtevanih kriterijuma za prihvatljivost alfa koeficijenata koji će zavisti od vremena primene unutar istraživanja. U skladu sa nalazima ovog istraživanja, a na osnovu vrednosti koje odgovaraju prvom tercilu, može se predložiti da kriterijum prihvatljive vrednosti alfa koeficijenata na početku studija za depresivne instrumente iznosi alfa $>0,6$ dok za merenja na kraju studija alfa $>0,75$.

Niže vrednosti alfa koeficijenata na početku studija tumačene su od strane nekih autora kao posledica ograničenja varijabiliteta skorova usled primene kriterijuma za uključenje ispitanika u istraživanje (Cunningham i sar., 2011). Sa protokom vremena u studijama dolazi do povećanja varijabiliteta kako na nivou ajtema, tako i na nivou ukupnih skorova, što redukuje mernu grešku i stvara uslove za povećanje zajedničke varijanse ajtema i skala (Sayer i sar., 1993). U ovom istraživanju potvrđena je povezanost povećanja vrednosti alfa koeficijenta u vremenu sa povećanjem vrednosti varijabiliteta skorova. Osim ovoga, povećanju interne konzistentnosti i korelacija između skala tokom vremena verovatno doprinosi i povećanje međuposmatračke pouzdanosti na nivou ajtema i ukupnih skorova, što takođe redukuje ukupnu mernu grešku.

Dok visoke vrednosti Cronbach koeficijenta alfa znače da svaki ajtem može da bude zamena za sve ajteme na skali, Loevingerin koeficijent daje opšti utisak o stepenu u kome svi ajtemi zajedno ispunjavaju Guttmanov kriterijum savršene skale (Bjerg Bendsen i sar., 1997). Preporučene vrednosti za Loevingerin koeficijent homogenosti su 0,4 i više, pri čemu su vrednosti 0,3 i više samo prihvatljive (Mokken, 1971). Objedinjene vrednosti Loevingerinog koeficijenta za HAMD skalu na prvom merenju su ispod praga prihvatljivosti i iznose 0,18, dok su na poslednjoj proceni 0,33 što je iznad praga prihvatljivosti. I za MADRS skalu objedinjene vrednosti Loevingerinog koeficijenta su na prvom merenju ispod praga prihvatljivosti (0,22), dok su na poslednjoj proceni iznad preporučenog praga (0,44). Kao što se vidi, u oba slučaja vrednosti koeficijenata pouzdanosti na poslednjoj proceni su više nego na prvoj. U oba slučaja, takođe, nije bilo moguće testirati prediktore zbog numeričkih ograničenja RVE metoda.

Za Guttmanov koeficijent dve polovine nije bilo moguće uraditi objedinjavanje vrednosti jer je saopšten samo u jednoj studiji sa ponovljenim merenjima za HAMD skalu (Gençöz i sar., 2007). U pomenutoj studiji vrednosti Guttmanovog koeficijenta dve polovine više su na poslednjoj proceni (0,85) nego na prvoj proceni (0,82).

Za međuposmatračku pouzdanost podaci sa ponovljenih procena su manje dostupni u odnosu na podatke o internoj konzistentnosti. Ovo je u suprotnosti sa nalazima meta-analize o pouzdanosti HAMD skale (Trajković i sar., 2011). Razlog je najverovatnije neophodno učešće najmanje dva ocenjivača za ocenu međuposmatračke pouzdanosti, koji pod sličnim uslovima ponovljeno ocenjuju ispitanika, što predstavlja izvestan organizacioni problem u samom istraživanju, ako već ocena međuposmatračke pouzdanosti nije primarni cilj, dok je za primenu ocena interne konzistentnosti dovoljan jedan ocenjivač. Neophodno učešće dva ocenjivača je takođe razlog što se međuposmatračka pouzdanost gotovo uvek ocenjuje na podskupu, a ne na čitavom uzorku ispitanika u istraživanju. To je u ovom istraživanju nametnulo potrebu uvođenja dodatnog kriterijuma za podatke o međuposmatračke pouzdanosti, a to je da je eksplicitno saopštena veličina uzorka na kojoj je ona ocenjena, što je sa svoje strane dodatno smanjilo broj raspoloživih podataka za analizu.

Od metoda za ocenu međuposmatračke pouzdanosti najčešće je primenjen metod intraklasnog koeficijenta korelacije. Ovo je u skladu sa preporukama u literaturi da za

izračunavanje koeficijenta međuposmatračke pouzdanosti treba dati prednost ICC-u u odnosu na Pearsonov i Spearmanov koeficijent korelacije koji mogu dati njegove precenjene vrednosti (Aday i Cornelius, 2006). Kao i u slučaju objedinjenih vrednosti alfa koeficijenta, i ovde je zapažen porast vrednosti ICC-a u vremenu. Objašnjenje može biti slično kao i u slučaju porasta alfa koeficijenta u vremenu, ograničenjem varijabiliteta skorova usled primene kriterijuma za uključanje ispitanika u istraživanje na početku studija, dok sa protokom vremena u studijama dolazi do povećanja varijabiliteta skorova, što za posledicu ima redukciju merne grešku. Osim ovoga, verovatno je da porastu međuposmatračke pouzdanosti doprinosi i porast veština samog posmatrača da bolje primenjuje instrumente i bolje opservira sve karakteristike ispitanika koje su zahtevane stavkama instrumenta. Ovo isključuje trening ocenjivača između dve procene, jer je kriterijum za uključanje u studiju bio nepromenjenost uslova u ponovljenim procenama. Primena treninga u toku same studije, sa ciljem unapređenja veština ocenjivanja, automatski znači promenu uslova ocenjivanja, pa takve studije nisu uključene u analizu. Međutim, može se pretpostaviti da prva procena i primena instrumenta predstavljaju neku vrstu treninga za ocenjivače, što rezultuje unapređenjem veštine ocenjivanja prilikom kasnijih procena.

Idealni podaci za analizu promena test-retest pouzdanosti u vremenu podrazumevali bi najmanje četiri procene, pri čemu bi prvi koeficijent test-retest pouzdanosti bio dobijen primenom koeficijenta korelacije između prve i druge procene, a ponovljeni koeficijent test-retest pouzdanosti kao koeficijent korelacije između treće i četvrte procene. Uslov da se to uopšte može smatrati koeficijentima test-retest pouzdanosti je približna nepromenjenost uslova merenja na ponovljenim procenama. Ovo podrazumeva da su na svim procenama ispitanici ocenjeni od strane istog ocenjivača.

Jedna od retkih preporuka, koja potiče iz literature o naučnom metodu u medicini i bihejvioralnim naukama, za prihvatljivu vrednosti koeficijenta test-retest pouzdanosti navodi je vrednost od 0,8 (Gliner i sar., 2009). Pitanje je da li se samo jedna predložena vrednost može generalno prihvatiti u medicini, uključujući i psihijatrijske skale. Naime, može se očekivati da su vrednosti test-retest pouzdanosti povezane sa varijabilitetom skorova, kao što je to slučaj za druga dva tipa pouzdanosti. Osim toga na vrednost test-retest pouzdanosti utiče i vreme proteklo između dva merenja. U meta-analizi pouzdanosti HAMD skale (Trajković i sar., 2011) nađena je negativna povezanost između

vremenskog intervala između dva intervjua i vrednosti koeficijenata test-retest pouzdanosti (Spearman $\rho = -0,74$). Interval između dva intervjua bio je u opsegu od 1 do 30 dana. Jasno je da prilikom ocene test-retest pouzdanosti vreme između ponovljenih procena ne sme biti preveliko da ne bi došlo do značajne promene kliničkog stanja ispitanika, odnosno do znatne promene stanja objekta merenja. Vreme između dve procene ne sme biti ni prekratko, zbog memorijskog efekta kako na ocenjivača tako i na ispitanika. Naime, sećanje na prvi intervju može imati efekat na ispitanika prilikom odgovaranja na postavljena pitanja na ponovljenom intervjuu, a istovremeno sećanje ocenjivača na prvi intervju može imati efekat na to kako će postavljati pitanja i kako ocenjivati ispitanika (Miles i Gilbert, 2005). Postojanje efekta sećanja na taj način za rezultat može imati slično ponašanje i ispitanika i ocenjivača na ponovljenim procenama, što nerealno redukuje mernu grešku i povećava ocenu pouzdanosti.

U ovom istraživanju nisu nađene studije u kojima je eksplicitno navedeno da se radi o ponovljenim ocenama test-retest pouzdanosti. Međutim, iskorišćeni su publikovani individualni skorovi na HAMD skali iz najmanje četiri različita vremena pod uslovom da su dobijeni od strane istog ocenjivača, što je detektovano u tri studije. Jedna studija je zadovoljila kriterijum da ima iste ocenjivače pri ponovljenim merenjima ali zbog velike razlike u vremenu merenja druge test-retest ocene ta studija nije uključena u meta-analizu. Objedinjene vrednosti test-retest pouzdanosti pokazuju povećanje u vremenu. Broj podataka nije dopustio da se metodama meta-regresije ispita povezanost ovih promena sa potencijalnim prediktorima, ali može se pretpostaviti da su i ove promene, kao i u slučaju alfa koeficijenata, pod uticajem povećanja varijabiliteta skorova i povećanja druga dva tipa pouzdanosti. Objedinjene vrednosti test-retest pouzdanosti su dovoljno visoke i ukazuju da ispitivani instrument (HAMD skala u ovom slučaju) pouzdano meri stanje depresivnosti, kao i da promene skorova pouzdano reflektuju promene stanja depresivnosti.

U meta-analizama koje ispituju promenu u korelacijama tokom vremena između tri instrumenta koji se koriste za ocenu težine depresivnih simptoma glavni rezultat je da se ove korelacije povećavaju tokom vremena u longitudinalnim studijama. Ovo povećanje nađeno je i za korelacije ocenjivačkih skala (HAMD i MADRS) i između jedne ocenjivačke skale (HAMD ili MADRS) i skale za samoprocenu (BDI). Štaviše, ova meta-analiza je po prvi put otkrila povećanje koeficijenata korelacije između skala MADRS i

BDI tokom vremena. Povećanje koeficijentata korelacija između tri depresivne skale, prikazano u sve tri meta-analize, je znatno; u poređenju sa početnim vrednostima, koeficijenti korelacije su za oko jednu trećinu veći na poslednjem ocenjivanju. Najveća promena nađena je za koeficijent korelacije između skala HAMD i BDI, verovatno jer su koeficijenti korelacije ovih skala na početku niži nego korelacije ostala dva para instrumenata.

U ovoj studiji najjače korelacije su nađene između skala koje procenjuju ocenjivači, tj. HAMD i MADRS, najverovatnije zato što dele određene stavke na skali, imaju sličan način primene i ocenjivanja, i generalno ocenjuju se od strane istog ocenjivača. Instrumenti koje procenjuju ocenjivači mogu biti podložni pristrasnosti ocenjivača, dok instrumenti samoprocene mogu biti podložni pristrasnosti samoocenjivača (Feliciano i Gum, 2009; Lin i sar., 2014). Takođe, postoje i razlike u sadržaju pomenute tri skale. HAMD skala procenjuje širok spektar simptoma sa fokusom na somatske simptome, dok MADRS skala procenjuje raspoloženje, somatske i kognitivne simptome, izuzev psihomotornih simptoma (retardacija i agitacija). Suprotno od toga, BDI skala uglavnom meri bihevioralne i kognitivne karakteristike depresije (Schrag i sar., 2007; Stefanis i Stefanis, 2003; Thompson, 1989). Pored toga karakteristike pacijenata mogu da utiču na korelacije između skala. Enns i sar. (2000) su evaluirali korelacije između skala procene od strane ocenjivača i skala samoprocene depresije i ustanovili da uzrok neslaganja mogu biti neke karakteristika pacijenata, na primer, stariji pacijenti imaju tendenciju da umanjuju simptome na BDI pa je skor niži u poređenju sa skalama koje procenjuju ocenjivači.

Podaci o korelacijama između skala za procenu depresije tokom longitudinalnih studija su mešoviti. Neki autori saopštavaju povećanje korelacija između skala za procenu depresije tokom vremena. Na primer, Sayer i sar. (1993) nalaze srednju korelaciju na početku ($r=0,48$) između skala HAMD i BDI; ova korelacija se povećala značajno u kasnijim merenjima ($r=0,78$). Slično, Brown i sar. (1995) navode da se koeficijenti korelacije između skala HAMD i BDI konzistentno povećavaju tokom vremena, u rasponu 0,70 nakon prvog meseca, do 0,85 nakon 8 meseci. Dalje, Zimmerman i sar. (2004) saopštavaju da je u tri studije došlo do povećanja korelacija između skala HAMD i MADRS tokom tretmana antidepresivima, a Jiang i Ahmed (2009) su našli povećanje korelacija između skala HAMD i MADRS tokom tretmana. Suprotno od ovih

istraživanja, El-Giamal i sar (2003) našli su smanjenje koeficijenta korelacije između skala HAMD i BDI, dok su Ramasubbu i sar. (2013) saopštili smanjenje koeficijenta korelacije između skala HAMD i MADRS tokom perioda posmatranja.

Najveći broj studija uključenih u ove meta-analize ima mali broj ispitanika, koji su regrutovani iz različitih populacija i ustanova. Ovo delimično može objasniti različite rezultate. U meta-regresiji, svi faktori koji su mogli biti izvor heterogenosti između uključenih studija su uključeni kao potencijalni prediktori korelacija između skala. Meta-regresija je pokazala da postoje dva prediktora povećanja korelacija između skala za procenu depresije u vremenu: smanjenje ukupnog skora i povećanje varijabiliteta skorova. Što se tiče prvog prediktora, postoji povezanost između smanjenja skora depresivnosti, tj. redukcije simptoma i poboljšanja kliničkih manifestacija depresije, i porasta korelacija između skala za procenu depresije. Ovo je u skladu sa prethodnim istraživanjima Daviesa i sar. (1975) da se više korelacije između skala dobijaju kod pacijenata sa blažim simptomima. Rezultati nekoliko drugih studija, takođe, sugerišu da su promene u korelacijama između skala za procenu depresije povezane sa promenama u težini depresivnih simptoma (Bailey i Coppen, 1976; Martinsen i sar., 1995; Seemüller i sar., 2012).

Druga istraživanja objašnjavaju porast korelacija između skala za procenu depresije kroz povećanje opsega depresivnih skorova (Alexopoulos i sar., 2014; Brown i sar., 1995; Flint i Rifat, 1996; Sayer i sar., 1993; Senra i Polaino, 1993; Zimmerman i sar., 2004). Teorijski gledano, na korelacije može uticati varijabilitet i opseg skorova: ograničenje opsega može smanjiti koeficijent korelacije (Schmidt i Hunter, 2014; Urbina, 2014). Kao što je navedeno, rezultati ove studije takođe sugerišu da je porast korelacija između skala za procenu depresije povezan sa povećanjem varijabiliteta depresivnih skorova. Razlike između ispitanika su obično manje na početku studije jer tada većina ispitanika ima više depresivne skorove. Ovo je naročito slučaj sa studijama koje su koristile postojanje kriterijuma za uključanje u studiju koji je u vezi sa težinom bolesti (tj. definisan kao minimalni skor depresivnosti). Tokom vremena, a naročito usled primene tretmana, varijabilitet depresivnih skorova postaje veći, jer neki ispitanici reaguju bolje na tretman nego ostali, što se reflektuje na njihov skor depresivnosti.

Brown i sar. (1995) komentarišu da niske inicijalne korelacije između skala za procenu depresije mogu biti izazvane korišćenjem kriterijuma za uključenje u studiju koji je u vezi sa težinom bolesti. Sa obzirom da primena kriterijuma smanjuje varijabilitet depresivnih skorova na početku i da je porast varijabiliteta depresivnih skorova povezan sa porastom korelacija između skala za procenu depresije, očekivalo se da će u meta-regresiji u ovom istraživanju, primena kriterijuma za uključenje u studiju koji je u vezi sa težinom bolesti biti statistički značajan, nezavisni prediktor promena korelacija između skala za procenu depresije. Činjenica da to nije bio slučaj može biti povezana sa mogućnošću da primena kriterijuma za uključenje u studiju koji je u vezi sa težinom bolesti nije saopštena u nekim radovima.

Sayer i sar. (1993) zaključuju da su relativno slabe početne korelacije između skala HAMD i BDI uglavnom zbog sklonosti pacijenata da ocene svoje simptome niže u poređenju sa procenom ocenjivača. Ponovljene samoprocene mogu poboljšati sposobnost pacijenata da prate svoje simptome, što dovodi do jačih korelacija sa skalama procene od strane ocenjivača. Ova pretpostavka nije potvrđena u meta-analizi zbog nedostatka podataka o uvidu pacijenata u svoju bolest/depresiju i njihovu sposobnost da saopšte depresivne simptome. Osim toga, u meta-analizi je nađeno da korelacije rastu i između dve skale procene od strane ocenjivača (HAMD i MADRS). Sve ovo ukazuje na to da paralelno sa pretpostavljenim poboljšanjem pacijenata da prate svoje simptome, ponovljene procene mogu takođe da dovedu do poboljšanja veštine ocenjivača u proceni depresivnih simptoma.

Opšte je prihvaćeno da je visoka pouzdanost skale preduslov za njenu valjanost, iako visoka pouzdanost nije dovoljna da bi skala bila valjana (Mitchell i Jolley, 2012). Nije bilo moguće analizirati koeficijente pouzdanosti kao prediktore korelacija između skala za procenu depresije jer je samo 5 radova istovremeno saopštilo i ponovljene koeficijente pouzdanosti. Smer promena koeficijenata valjanosti u vremenu praćen je promenama koeficijenata pouzdanosti u istom smeru kod svih 5 radova. Prethodna meta-analiza pouzdanosti HAMD skale (Trajković i sar., 2011) saopštila je da su više vrednosti Cronbach alfa koeficijenta povezane sa nižim ukupnim skorovima i višim varijabilitetom skorova na HAMD skali. Na osnovu saopštenih vrednosti alfa koeficijenta u toj meta-analizi, došlo je do porasta alfa koeficijenata za HAMD skalu u 12 od 14 studija koje su saopštile ponovljene koeficijente pouzdanosti. To sugeriše da porast interne

konzistencije, a samim tim i porast korelacija između stavki na skali u longitudinalnim studijama može biti povezano sa porastom korelacija između skala mereći isti konstrukt. Drugim rečima, i pouzdanost i valjanost skala može se poboljšati tokom vremena.

Nekoliko analiza senzitivnosti potvrdile su zadovoljavajuću robusnost dobijenih rezultata. Prvo, skoro identični rezultati dobijeni su u dve istovremeno izvedene meta-regresione analize sa različitim zavisnom varijablama (korelacija između skala HAMD i BDI i korelacija između skala HAMD i MADRS). Drugo, analiza senzitivnosti izvedena je sa različitim vrednostima ρ u RVE metodi. Treće, analiza je ponovljena nakon isključenja 4 studije koje su koristile LOCF metod za prikupljanje podataka. U ostalim analizama, isključeno je 11 studija koje su koristile HAMD skale sa više od 17 stavki i devet studija koje nisu koristile originalni BDI. U sledećoj analizi, isključeno je 20 studija koje su male bar jednu korelaciju koja nije bila statistički značajna. Četvrto, koeficijenti korelacije između skala HAMD i BDI saopšteni su u radovima (22 studije) ili izračunati sekundarnom analizom iz IPD (22 studije); rezultati odvojenih meta-analiza publikovanih ili izračunatih koeficijenata korelacije su isti kao oni dobijeni iz meta-analize svih koeficijenata korelacije. Poređenje publikovanih i izračunatih koeficijenata korelacije ukazuje da nema publikacione pristrasnosti. Studije koje su saopštile IPD imaju manju veličinu uzorka, ali one takođe pružaju više podataka za moguće prediktore i na taj način su zastupljenije u meta-regresionim modelima. Peto, nije nađena povezanost između smanjenja veličine uzorka tokom vremena i korelacija između skala.

Upotreba validnih instrumenata za procenu depresije važna je za kvantifikovanje promena u stepenu depresije tokom vremena. Ispitivanje promena korelacija između skala u toku vremena pomaže da se utvrdi konkurentna validnost ovih instrumenata u različitim tačkama u vremenu. Uzimajući u obzir rezultate ove meta-analize i drugih studija koje su saopštile promene korelacija između skala za procenu depresije u vremenu, neophodno je uspostaviti prihvatljive koeficijente korelacija za longitudinalne studije, koje bi ukazivale na konkurentnu valjanost ovih skala.

Za razliku od koeficijenata pouzdanosti, ne postoji opšti konsenzus za vrednosti koeficijenta korelacije koje ukazuju na prihvatljivu valjanost. Neki autori sugerišu $r=0,3-0,4$ (Muchinsky, 2006) kao prihvatljiv prag za korelacije između skala koje mere isti konstrukt, dok drugi sugerišu $r=0,6$ ili više (McDowell, 2006). Koeficijenti korelacije

iznad 0,9 smatraju se poželjnim a koeficijenti iznad $r=0,8$ prihvatljivim za druge autore (Safrit i Wood, 1995). Preporuke o prihvatljivim korelacijama između skala za procenu depresije su takođe nekonzistentne. Bagby i sar. (2004) sugerišu da korelacije iznad 0,5 bi mogle da ukazuju na prihvatljivu konkurentnu valjanost. Suprotno od njega, Scogin i sar. (1988) ne smatraju da koeficijenti korelacije od 0,42 i 0,54 između skala HAMD i BDI ukazuju na prihvatljivu konkurentnu valjanost.

Nedostatak opštih utvrđenih kriterijuma za prihvatljive koeficijente korelacija između skala, razlike između instrumenata za procenu depresije i nalaz da korelacije između skala za procenu depresije rastu tokom vremena ukazuju na to da su neophodni različiti pragovi adekvatnih korelacija na početku i na kraju longitudinalnih studija. Rezultati dobijeni u ovom istraživanju sugerišu sledeće pragove, a oni su bazirani na osnovu vrednosti koje odgovaraju prvom tercilu, tj. vrednost koja seče prvu trećinu distribucije empirijskih učestalosti ekstrahovanih koeficijenata korelacije:

1. Za korelacije između skala procene od strane ocenjivača i skala samoprocene: 0,4 na početku i 0,7 na kraju studije;
2. Za korelacije između dve skale procene od strane ocenjivača: 0,65 na početku i 0,85 na kraju studije.

Meta-analizu promena dijagnostičkih performansi u vremenu izvedena je sa ciljem analize promene dijagnostičke tačnosti, odnosno dijagnostičke valjanosti psihijatrijskih skala u vremenu. Psihijatrijske dijagnoze se postavljaju primenom dijagnostičkih kriterijuma, kao što je DSM-IV, ICD-10 ili RDC (Dittmann, 1991; Stein i sar., 2013; Türkçapar i sar., 1999). Rezultat primene dijagnostičkih kriterijuma je dihotomija (binarnost) u smislu ima ili nema određene dijagnoze. Rezultat primene psihijatrijskih skala su ukupni skorovi sa osobinama koje su približne osobinama numeričkih kontinuiranih podataka. Konvertovanjem takvih podataka u dihotomne, u cilju postavljanja dijagnoze, gubi se značajan deo informativnosti ukupnih skorova. Osim toga, zlatni standard u psihijatriji, odnosno dijagnostički kriterijumi, nemaju nijedan nedostatak, kao što je to čest slučaj u medicini, pa i nema pravog razloga da se umesto njih u tu svrhu koriste psihijatrijske skale. I pored toga, u nekim studijama je ispitana i ta osobina psihijatrijskih skala. Takve studije su relativne retke, a naročito retki su takvi

podaci o dijagnostičkim performansama publikovani za više od jednog merenja u longitudinalnim studijama. To je i očekivano jer psihijatrijske skale nisu namenjene postavljanju dijagnoze već praćenju promena težine kliničkog stanja u vremenu. U medicini ponovljena primena dijagnostičkih testova je u cilju povećanja dijagnostičkih performansi (Lau, 1989; Wang i Geisser, 2003; Zou, 2001). Osim ovoga, za psihijatrijske skale je od značaja ponovljena ocena dijagnostičkih performansi da bi se utvrdila njihova merna adekvatnost u uslovima kada se očekuje izlečenje ili nastupanje remisije.

U ovom istraživanju pokazano je da su promene ocena dijagnostičke adekvatnosti male. Kod obe ispitivane skale (HAMD i MADRS), dolazi do povećanja dijagnostičke efektivnosti (DOR) kao i do povećanja specifičnosti. U slučaju HAMD skale dolazi do malog pada senzitivnosti, dok se u slučaju MADRS skale ova mera tačnosti ne menja. Povećanje dijagnostičke efektivnosti u vremenu može se objasniti, jednim delom, povećanjem pouzdanosti koja je nađena u ovom istraživanju. Kako je pouzdanost preduslov valjanosti, odnosno preciznost preduslov tačnosti, može se i očekivati da će porast pouzdanosti psihijatrijskih skala u vremenu biti povezan i sa porastom dijagnostičke tačnosti. Moguće objašnjenje za povećanje specifičnosti za skale HAMD i MADRS, kao i smanjenja senzitivnosti za HAMD skalu, je pristrasnost koja se javlja kod ocenjivača na kraju perioda praćenja i tendencije da težinu kliničkog stanja ocene blažom nego što jeste. Povećanje specifičnosti u vremenu je od značaja, jer sa protokom vremena se može očekivati pad prevalencije depresije u svakom uzorku, usled izlečenja ili remisije pacijenata, a ujedno se može očekivati i porast stope lažno pozitivnih. Povećanje specifičnosti u vremenu može bar delimično ograničiti porast stope lažno pozitivnih.

Faktorska valjanost je poseban tip konstruktivne valjanosti koji se odnosi na korelisanost ajtema unutar skale. Faktorska valjanost se ispituje primenom faktorske analize (Myers i Winters 2002). Rezultat faktorske analize jesu faktori koji se dobijaju linearnom kombinaciju ajtema, i koji predstavljaju dimenzije odnosa ajtema u sastavu skale. Varijable sa najvećim opterećenjem na nekom faktoru od najvećeg su značaja za interpretaciju tog faktora. Ako se za neku skalu očekuje da je jednodimenzionalna i da meri samo jedan konstrukt idealno rešenje je da se u faktorskoj analizi dobije samo jedan faktor, a za skale koje se sastoje od podskala više faktora (Fajgelj 2003; Myers i Winters 2002). Jednodimenzionalnost se lakše postiže skalama sa manje ajtema, dok duže skale, koje obično osim što mere ciljani konstrukt mere delom i srodne konstrukte.

Rezultati faktorskih analiza, u smislu broja ekstrahovanih faktora, objašnjene varijanse i strukture faktora, razlikuju se od studije do studije. U faktorskim analizama HAMD skale ukazano je na nepostojanje stabilne faktorske strukture. Broj ekstrahovanih faktora varira od dva do šest (Möller, 2014). U originalnoj faktorskoj strukturi HAMD skale, opisanoj od strane Hamiltona, opisana su tri faktora: opšti faktor depresivne bolesti, bipolarni sa simptomima anksioznosti (psihičke i somatske) i agitacije (Hamilton, 1967, 1960). U literaturi je prvi faktor najčešće označen kao melanholični i obuhvata ajteme označene kao jezgro depresivnih simptoma: depresivno raspoloženje, osećanje krivice, rad i aktivnost, retardaciju, anksioznost, i opšte telesni simptomi. Drugi faktor je često označen kao anksioznost i obuhvata ajteme agitacija, anksioznost i telesna anksioznost, dok je treći faktor često označen kao insomnija i obuhvata ajteme rana insomnija, teškoće prosnivanja i kasna insomnija (Bech, 2012; Fleck i sar., 1995). Zbog toga, neki autori smatraju da HAMD skala ne meri samo konstrukt depresije već i anksioznost kao i insomniju (Fiske i O'Riley, 2008). Faktorska struktura nemačke verzija HAMD-a poseduje osobinu unidimenzionalnosti kako na prvom tako i na ponovljenom merenju, ali i pored toga autori nisu bili u mogućnosti da jasno definišu opšti faktor (Maier i sar., 1985).

Prethodna istraživanja ukazuju da iako MADRS skala nije jednodimenzionalna skala, njen prvi faktor uključuje više ajtema nego što je to slučaj sa HAMD skalom. MADRS skala sadrži manje somatskih ajtema u odnosu na HAMD, i ima faktorsku strukturu koja je bliža merenju aspekata konstrukta depresije kao što su disforija i anhedonija. U faktorskim analizama MADRS skale dimenzije se mogu klasifikovati pod nazivima tužne/pesimističke misli, unutrašnje tenzije, nesposobnost osećanja i redukovan apetit (Möller, 2014).

Objašnjena varijansa na prvom ekstrahovanom faktoru, kao i ukupna objašnjena varijansa, u faktorskoj analizi skala HAMD i MADRS je u svih devet studija veća na drugom merenju u odnosu na prvo merenje, ali zbog malog broja studija razlika je na granici statističke značajnosti u oba slučaja. Povećanje objašnjene varijanse, kako na prvom ekstrahovanom faktoru tako i ukupne objašnjene varijanse, u faktorskim analizama na ponovljenim merenjima moglo bi se objasniti povećanjem pouzdanosti za skale HAMD i MADRS na ponovljenim merenjima koje je nađeno u ovom istraživanju. Povećanje pouzdanosti na ponovljenim merenjima podrazumeva povećanje korelacije

između ajtema što rezultuje povećanjem objašnjene varijanse, kako ukupne tako i na prvom ekstrahovanom faktoru.

Broj ekstrahovanih faktora na ponovljenim merenjima HAMD skale se ne menja. Ovo ukazuje, da i pored povećanja interne konzistentnosti u vremenu, povećanja korelacije između ajtema i objašnjene varijanse u faktorskim analizama, broj faktora HAMD skale pokazuje osobine vremenske invarijantnosti, i sličnost faktorske strukture na prvom i ponovljenim merenjima, odnosno da je HAMD skala multidimenzionalan instrument koji meri više od jednog konstrukta.

U faktorskim analizama MADRS skale smanjenje broja faktora na drugom merenju je na granici statističke značajnosti. U studiji Möller i Schnitker (2007) ukazano je da smanjenje broja faktora MADRS skale na drugom merenju verovatno nastaje zbog promene skorova na MADRS skali i smanjenja težine depresije, kao i povećanja sposobnosti ocenjivača da bolje ocenjuju depresivnost. U ovom istraživanju pokazano je da u ponovljenim merenjima dolazi do porasta pouzdanosti (interna konzistenost i međuposmatračka pouzdanosti) što takođe mogući uzrok smanjenja broja faktora, koji se ovde, za razliku od HAMD skale, ispoljio zbog veće fokusiranosti MADRS skale na merenje konstrukta depresije.

Longitudinalne studije su dominantan tip kliničkih studija koje se koriste u psihijatriji. Postoje različite definicije i tumačenja longitudinalnih studija. U studijama sa dizajnom ponovljenih merenja, iste varijable se mere više nego jednom, na datom uzorku jedinica. Postoji više termina u literaturi: dizajn ponovljenih merenja, pretest/posttest dizajn, longitudinalni dizajn, crossover dizajn (Stommel i Wills, 2004). Longitudinalna studija je ona koja se odvija tokom vremena, postoje najmanje dva (a često i više) merenja na istim studijskim jedinicama. Razlikuju se dva tipa longitudinalnog dizajna: ponovljena merenja i vremenske serije, koja se razlikuju na osnovu broja merenja tokom vremena (20 je konvencionalno korišćena granica) (Kirch, 2008). Klinički ogledi su po definiciji longitudinalne studije. To su na slučajan način uključeni pacijenti na različite tretmane, pacijenti se prate tokom vremena, prate se ishodi i na kraju porede grupe prema tretmanu (Walters, 2009). U longitudinalnim studijama ispitanici se prate rokom vremena sa kontinuiranim ili ponovljenim praćenjem faktora rizika ili zdravstvenih ishoda, ili oboje (Coggon i sar., 2009). Za studiju se kaže da je longitudinalna ili studija praćenja, kada se

ispitanici prate od ulaska u studiju do utvrđivanja određenog ishoda od interesa, gubitka u periodu praćenja ili administrativnog kraja praćenja, šta god se prvo desi (Toh i Hernán, 2008). U ovom istraživanju pod longitudinalnim studijama podrazumevale su se opservacione ili eksperimentalne studije sa ponovljenim merenjima istih varijabli na istim ispitanicima u različitim tačkama u vremenu.

Ovo je prvo istraživanje promena koeficijenata pouzdanosti i valjanosti za psihijatrijske instrumente u longitudinalnim studijama. U metodološkom smislu novina je primena meta-regresije sa robusnom ocenom varijanse u procesu modelovanja koeficijenata pouzdanosti i valjanosti u vremenu. Sveobuhvatno su dati grafički prikazi za ovaj tip podataka, što je uključivalo objedinjene regresione nagibe, sukcesivni prikaz forest plotova u dva vremena, publikaciona pristrasnost i heterogenost podataka. Na osnovu pregleda dostupne stručne literature, ovo je prvi put da su dobijene ocene test-retest pouzdanosti u longitudinalnim studijama, koje su u aktuelnom istraživanju dobijene na osnovu publikovanih individualnih podataka pacijenata. Po prvi put u literaturi, u ovom istraživanju, na osnovu dovoljnog broja podataka bilo je moguće generisati preporuke za pragove prihvatljivih vrednosti alfa koeficijenta i koeficijenata korelacije na prvoj i poslednjoj proceni u longitudinalnim studijama.

Glavni nedostatak ovog istraživanja je da meta-regresioni modeli nisu izvedeni za sve ocene pouzdanosti i valjanosti. Univarijantna meta-regresija izvedena je za ocene: alfa koeficijenta, Loevingerinog koeficijenta i ICC za HAMD skalu, i za korelacije između skala HAMD i BDI, i HAMD i MADRS. Multivarijantna meta-regresiona analiza je jedino izvedena za ocene (1) Alfa koeficijenta za HAMD skalu i (2) korelacije između skala HAMD i BDI. U drugim slučajevima, meta-regresioni modeli nisu izvedeni iz više razloga: (1) numerička ograničenja za meta-regresionu analizu sa RVE gde stepeni slobode postaju manji od 4, (2) uključivanje više od jednog prediktora u model dovodi do smanjenja broja studija zbog nedostajućih podataka, i (3) nepovoljnog odnosa između broja prediktora i broja studija. Druga ograničenja su da nisu uzeti u obzir primenjeni lekovi i tretmani tokom trajanja studija kao nezavisni prediktori u meta-regresiji zbog nedostatka podataka.

Postojećim statističkim metodama adekvatno se može ispitati promena pouzdanosti i valjanosti psihijatrijskih instrumenata u vremenu u longitudinalnim studijama. Meta-analitičkim metodom robusne ocene varijanse modelovane su promene koeficijenata pouzdanosti i konkurentne valjanosti u vremenu, kao i povezanost tih promena sa potencijalnim prediktorima, dok su metodom recipročne varijanse objedinjeni koeficijenti pouzdanosti i konkurentne valjanosti u pojedinačnim vremenima.

U studiji slučaja, u ovom istraživanju, za HAMD skalu nađen je porast interne konzistentnosti i međuposmatračke pouzdanosti. Mali broj podataka nije mogao dovesti do zaključka o promenama test-retest pouzdanosti HAMD skale u vremenu, iako bi se zbog povećanja prethodna dva tipa pouzdanosti očekivalo da to i ovde bude slučaj. Takođe bi se očekivalo da takve promene budu nađene i za MADRS skalu, ali i u ovom slučaju mali broj podataka vodio je u inkonkluzivnost o promenama interne konzistentnosti i međuposmatračke pouzdanosti, dok podaci o test-retest pouzdanosti nisu bili dostupni. To je takođe bio slučaj i za dijagnostičku i faktorsku valjanost, kako za HAMD skalu, tako i za MADRS.

Iz navedeno se može zaključiti da je mali broj publikovanih podataka u ovom tipu istraživanja glavna prepreka za dostizanje opšteg zaključka o promenama pouzdanosti i valjanosti psihijatrijskih skala u vremenu u longitudinalnim studijama. Moguća objašnjenja zbog čega su ti podaci retki jesu: (a) nedovoljna zainteresovanost istraživača da ispitaju merne karakteristike instrumenata u svom istraživanju, (b) organizacioni problemi ponovljenog angažovanja ocenjivača, i to najmanje dva puta po dva ocenjivača za ponovljenu međuposmatračku pouzdanost, kao i najmanje četiri puta za istog ocenjivača za ponovljenu test-retest pouzdanost, (c) neke mere pouzdanosti i valjanosti, npr. faktorska valjanost, su komplikovane za interpretaciju, a ponovljeno merenje interpretaciju čini još težom, i (c) nedovoljnog znanja istraživača da su moguće promene merne adekvatnost instrumenata u vremenu.

Mali broj publikovanih podataka o ponovljenim ocenama pouzdanosti i valjanosti u longitudinalnim studijama delom je nadomešten korišćenjem publikovanih individualnih podataka. Na osnovu reanalize IPD-a dobijeno je 50% podataka za meta-analizu promene u vremenu koeficijenata korelacije između skala HAMD, MADRS i BDI. Takođe, reanalizom IPD-a dobijeni su jedini dostupni podaci o ponovljenoj test-

retest pouzdanosti (dve studije sa publikovani individualnim skorovima na HAMD skali dobijenih od strane istog istraživača). Može se zaključiti da je reanaliza IPD-a alternativni metod za povećanje brojnosti podataka o korelacijama između skala i test-retest pouzdanosti u ponovljenim merenjima, ili njihove dostupnosti ako takvi podaci nisu ni publikovani.

U ovom istraživanju jedino je dostupnost podataka o koeficijentu alfa za HAMD skalu i korelacijama između ispitivanih instrumenata bila takva da je bilo moguće na osnovu empirijske raspodele dati preporuke za prihvatljive vrednosti koje su specifične za vreme primene.

Ovo istraživanje, i pored toga što nije pružilo osnovu za dostizanje opšteg zaključka o promenama pouzdanosti i valjanosti psihijatrijskih instrumenata u vremenu u longitudinalnim studijama, već samo parcijalne nalaze o povećanju interne konzistentnosti, međuposmatračke pouzdanosti i korelacije između ispitivanih skala, ukazuje na značaj moguće promene merne adekvatnosti psihijatrijskih instrumenata u vremenu. Rezultati i zaključci većine psihijatrijskih istraživanja bazirana su na primeni najmanje jednog instrumenta koji podrazumeva upotrebu ukupnih skorova ili skorova sa individualnih ajtema. Mnoge od tih studija su longitudinalne, pa je od velikog značaja ispitivanje merne adekvatnosti psihijatrijskih instrumenata u ponovljenim merenjima.

6. Zaključci

Zaključci koji se odnose na meta-analitički metod ispitivanja promena pouzdanosti i valjanosti psihijatrijskih skala u vremenu su:

1. Postojećim meta-analitičkim metodama adekvatno se mogu oceniti pokazatelji pouzdanosti i valjanosti psihijatrijskih instrumenata u longitudinalnim studijama:
 - a. Metodom robusne ocene varijanse analiza promene koeficijenata pouzdanosti i konkuretna valjanosti u vremenu, kao i povezanost tih promena sa potencijalnim prediktorima.
 - b. Metodom recipročne varijanse objedinjavanje koeficijenata pouzdanosti i konkuretna valjanosti u pojedinačnim vremenima.
 - c. Sekundarna analiza publikovanih individualnih podataka dopušta izračunavanje koeficijenata korelacija između skala, kao i test-retest pouzdanosti (pod uslovom istovetnosti posmatrača). To je alternativni metod za povećanje brojnosti podataka o korelacijama između skala ili njihove test-retest pouzdanosti u ponovljenim merenjima, ili njihove dostupnosti ako takvi podaci nisu ni publikovani.
2. Meta-analiza ocena pouzdanosti i valjanosti u longitudinalnim studijama pruža objektivnu osnovu za generisanje preporuka njihove prihvatljive vrednosti koje su specifične za instrument/oblast i vreme primene.
3. Niska učestalost saopštavanja podataka o pouzdanosti i valjanosti primenjenih psihijatrijskih instrumenata predstavlja glavni problem za efektivno korišćenje meta-analitičkog metoda u istraživanju promena merne adekvatnosti u longitudinalnim studijama.

Zaključci koji se odnose na meta-analizu promena ocena pouzdanosti i valjanosti za Hamiltonovu skalu za procenu depresivnosti i Montgomeri-Ašbergovu skalu za procenu depresivnosti:

1. Pouzdanost Hamiltonove skale za procenu depresivnosti, interna konzistenost, iskazana u vidu Cronbachovih alfa koeficijenata i Loevingerinih koeficijenata, kao i međuposmatračka pouzdanost, iskazana u vidu intraklasnih koeficijenata korelacije, pokazuje povećanje u vremenu u longitudinalnim studijama.
2. Koeficijenti konkurentne valjanosti Hamiltonove skale za procenu depresivnosti i Montgomeri-Ašbergove skale za procenu depresivnosti iskazani u vidu koeficijenata korelacija, kao i koeficijenti korelacije sa Bekovom skalom za procenu depresivnosti pokazuju trend povećanja u longitudinalnim studijama. Njihove vrednosti su za oko trećinu veće na kraju studija u odnosu na početak.

7. Literatura

1. Aday, L.A., Cornelius, L.J., 2006. *Designing and Conducting Health Surveys: A Comprehensive Guide*. John Wiley & Sons, San Francisco, CA.
2. Addington, D., Addington, J., Atkinson, M., 1996. A psychometric comparison of the Calgary Depression Scale for Schizophrenia and the Hamilton Depression Rating Scale. *Schizophr. Res.* 19, 205–212.
3. Addington, D., Addington, J., Schissel, B., 1990. A depression rating scale for schizophrenics. *Schizophr. Res.* 3, 247–251.
4. Agresti, A., 2010. *Analysis of Ordinal Categorical Data*. John Wileys & Sons, New York.
5. Alexopoulos, G., Kasper, S., Möller, H.-J., Moreno, C., 2014. *Guide to Assessment Scales in Major Depressive Disorder*. Springer, Heidelberg.
6. Arean, P.A., Perri, M.G., Nezu, A.M., Schein, R.L., Christopher, F., Joseph, T.X., 1993. Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. *J. Consult. Clin. Psychol.* 61, 1003–10.
7. Asberg, M., Montgomery, S.A., Perris, C., Schalling, D., Sedvall, G., 1978. A comprehensive psychopathological rating scale. *Acta Psychiatr. Scand. Suppl.* 5–27.
8. Baeken, C., De Raedt, R., Van Hove, C., Clerinx, P., De Mey, J., Bossuyt, A., 2009. HF-rTMS treatment in medication-resistant melancholic depression: results from 18FDG-PET brain imaging. *CNS Spectr.* 14, 439–48.
9. Baer, L., Blais, M.A., 2009. *Handbook of Clinical Rating Scales and Assessment in Psychiatry and Mental Health*.
10. Bagby, R.M., Ryder, A.G., Schuller, D.R., Marshall, M.B., 2004. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am. J. Psychiatry* 161, 2163–77.
11. Bailey, J., Coppen, A., 1976. A comparison between the Hamilton Rating Scale and the Beck Inventory in the measurement of depression. *Br. J. Psychiatry* 128, 486–9.
12. Bangert-Drowns, R.L., 1986. Review of developments in meta-analytic method. *Psychol. Bull.* 99, 388–399.
13. Barefoot, J.C., Burg, M.M., Carney, R.M., Cornell, C.E., Czajkowski, S.M., Freedland, K.E., Hosking, J.D., Khatri, P., Pitula, C.R., Sheps, D., 2003. Aspects of social support associated with depression at hospitalization and follow-up assessment among cardiac patients. *J. Cardiopulm. Rehabil.* 23, 404–412.

14. Bartko, J.J., 1966. The intraclass correlation coefficient as a measure of reliability. *Psychol. Rep.* 19, 3–11.
15. Bauer, M., Pretorius, H.W., Constant, E.L., Earley, W.R., Szamosi, J., Brecher, M., 2009. Extended-release quetiapine as adjunct to an antidepressant in patients with major depressive disorder: results of a randomized, placebo-controlled, double-blind study. *J. Clin. Psychiatry* 70, 540–9.
16. Baujat, B., Mahé, C., Pignon, J.-P., Hill, C., 2002. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. *Stat. Med.* 21, 2641–52.
17. Baumann, U., 1976. [Methodologic studies of the Hamilton rating scale for depression (author's transl)]. *Arch. für Psychiatr. und Nervenkrankheiten* 222, 359–75.
18. Bech, P., 2012. Quality of Life and Rating Scales of Depression, in: Preskorn, S.H., Feighner, J.P., Stanga, C.Y., Ross, R. (Eds.), *Antidepressants: Past, Present and Future*. Springer Science & Business Media, Heidelberg.
19. Bech, P., Fava, M., Trivedi, M.H., Wisniewski, S.R., Rush, A.J., 2011. Factor structure and dimensionality of the two depression scales in STAR*D using level 1 datasets. *J. Affect. Disord.* 132, 396–400.
20. Bech, P., Stage, K.B., Nair, N.P., Larsen, J.K., Kragh-Sørensen, P., Gjerris, A., 1997. The Major Depression Rating Scale (MDS). Inter-rater reliability and validity across different settings in randomized moclobemide trials. Danish University Antidepressant Group. *J. Affect. Disord.* 42, 39–48.
21. Bech, P., Tanghøj, P., Cialdella, P., Andersen, H.F., Pedersen, A.G., 2004. Escitalopram dose-response revisited: an alternative psychometric approach to evaluate clinical effects of escitalopram compared to citalopram and placebo in patients with major depression. *Int. J. Neuropsychopharmacol.* 7, 283–90.
22. Bech, P., Wilson, P., Wessel, T., Lunde, M., Fava, M., 2009. A validation analysis of two self-reported HAM-D6 versions. *Acta Psychiatr. Scand.* 119, 298–303.
23. Beck, A.T., Beck, R.W., 1972. Screening depressed patients in family practice. A rapid technic. *Postgrad. Med.* 52, 81–5.
24. Beck, A.T., Rush, A.J., Shaw, B.F., Emery, G., 1979. *Cognitive therapy of depression*. Guilford Press, New York.
25. Beck, A.T., Steer, R.A., Brown, G.K., 1996. *Manual for the Beck Depression Inventory-II*. Psychological Corporation, San Antonio, TX.
26. Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J., 1961. An inventory for measuring depression. *Arch. Gen. Psychiatry* 4, 561–71.

27. Bent-Hansen, J., Lunde, M., Klysner, R., Andersen, M., Tanghøj, P., Solstad, K., Bech, P., 2003. The validity of the depression rating scales in discriminating between citalopram and placebo in depression recurrence in the maintenance therapy of elderly unipolar patients with major depression. *Pharmacopsychiatry* 36, 313–6.
28. Berg, A., Lönnqvist, J., Palomäki, H., Kaste, M., 2009. Assessment of depression after stroke: a comparison of different screening instruments. *Stroke*. 40, 523–9.
29. Bergman, J., Miodownik, C., Palatnik, A., Lerner, V., 2011. Efficacy of Bupropion XR in Treatment-Resistant Elderly Patients. *Clin. Neuropharmacol.* 34, 17–20.
30. Bjerg Bendsen, B., Bjerg Bendsen, E., Lauritzen, L., Vilmar, T., Bech, P., 1997. Post-stroke patients in rehabilitation. The relationship between biological impairment (CT scanning), physical disability and clinical depression. *Eur. Psychiatry* 12, 399–404.
31. Blacker, D., 2009. Psychiatric Rating Scales, in: Sadock, B., Sadock, V., Ruiz, P. (Eds.), *Kaplan and Sadock's Comprehensive Textbook of Psychiatry*. Wolters Kluwer Health, Philadelphia, pp. 1032–1058.
32. Bland, J.M., Altman, D.G., 1997. Cronbach's alpha. *BMJ* 314, 572.
33. Bondolfi, G., Jermann, F., Rouget, B.W., Gex-Fabry, M., McQuillan, A., Dupont-Willemin, A., Aubry, J.-M., Nguyen, C., 2010. Self- and clinician-rated Montgomery-Asberg Depression Rating Scale: evaluation in clinical practice. *J. Affect. Disord.* 121, 268–72.
34. Borenstein, M., Hedges, L. V., Higgins, J.P.T., Rothstein, H.R., 2011. *Introduction to Meta-Analysis*. John Wiley & Sons.
35. Botella, J., Suero, M., Gambará, H., 2010. Psychometric inferences from a meta-analysis of reliability and internal consistency coefficients. *Psychol. Methods* 15, 386–97.
36. Bowen, R.C., 1978. The effect of diazepam on the recovery of endogenously depressed patients. *J. Clin. Pharmacol.* 18, 280–4.
37. Brakemeier, E.-L., Berman, R., Prudic, J., Zwillenberg, K., Sackeim, H.A., 2011. Self-evaluation of the cognitive effects of electroconvulsive therapy. *J. ECT* 27, 59–66.
38. Brown, C., Schulberg, H.C., Madonia, M.J., 1995. Assessment depression in primary care practice with the Beck Depression Inventory and the Hamilton Rating Scale for Depression. *Psychol. Assess.* 7, 59–65.
39. Brown, W., 1910. Some experimental results in the correlation of mental abilities. *Br. J. Psychol.* 1904-1920 3, 296–322.

40. Brunoni, A.R., Ferrucci, R., Bortolomasi, M., Vergari, M., Tadini, L., Boggio, P.S., Giacopuzzi, M., Barbieri, S., Priori, A., 2011. Transcranial direct current stimulation (tDCS) in unipolar vs. bipolar depressive disorder. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 35, 96–101.
41. Burrows, G.D., Davies, B., Scoggins, B.A., 1972. Plasma concentration of nortriptyline and clinical response in depressive illness. *Lancet* 2, 619–23.
42. Carneiro, A.M., Fernandes, F., Moreno, R.A., 2015. Hamilton depression rating scale and montgomery-asberg depression rating scale in depressed and bipolar I patients: psychometric properties in a Brazilian sample. *Health Qual. Life Outcomes* 13, 42.
43. Castonguay, L.G., Schut, A.J., Aikins, D.E., Constantino, M.J., Laurenceau, J.-P., Bologh, L., Burns, D.D., 2004. Integrative Cognitive Therapy for Depression: A Preliminary Investigation. *J. Psychother. Integr.* 14, 4–20.
44. Chabrol, H., Teissedre, F., Saint-Jean, M., Teisseyre, N., Rogé, B., Mullet, E., 2002. Prevention and treatment of post-partum depression: a controlled randomized study on women at risk. *Psychol. Med.* 32, 1039–47.
45. Cicchetti, D. V, Prusoff, B.A., 1983. Reliability of depression and associated clinical symptoms. *Arch. Gen. Psychiatry* 40, 987–90.
46. Cochran, W.G., 1954. The combination of estimates from different experiments. *Biometrics* 10, 101–129.
47. Coggon, D., Barker, D., Rose, G., 2009. *Epidemiology for the Uninitiated*.
48. Cohen, R.J., Swerdlik, M., 2009. *Psychological Testing and Assessment: An Introduction to Tests and Measurement*. McGraw-Hill Humanities.
49. Collet, L., Cottraux, J., Ladouceur, R., 1987. Cognitive therapy of depression and counterdemand effects: a pilot study. *Psychol. Rep.* 60, 555–60.
50. Cook, B.L., Helms, P.M., Smith, R.E., Tsai, M., 1986. Unipolar depression in the elderly. Reoccurrence on discontinuation of tricyclic antidepressants. *J. Affect. Disord.* 10, 91–4.
51. Cook, D.A., Beckman, T.J., 2006. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am. J. Med.* 119, 166.e7–16.
52. Cooper, H., Hedges, L. V., Valentine, J.C., 2009. *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation.
53. Cooper Harris M., 1998. *Synthesizing Research : A Guide for Literature Reviews Applied Social Research Methods Series*. Sage Publications, Inc.
54. Corn, T.H., Honig, A., Thompson, C., Bridges, P.K., Bartlett, J.R., Checkley, S.A., 1984. A neuroendocrine study of stereotactic sub-caudate tractotomy. *Br. J. Psychiatry* 144, 417–20.

55. Corruble, E., Sabran, V., Payan, C., Puech, A., Fermanian, J., Guelfi, J., 1998. Rating antidepressant efficacy with naturalistic live versus structured videotaped interviews. *Eur. Psychiatry* 13, 225–30.
56. Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334.
57. Cuijpers, P., Andersson, G., Donker, T., van Straten, A., 2011. Psychological treatment of depression: results of a series of meta-analyses. *Nord. J. Psychiatry* 65, 354–64.
58. Cuijpers, P., van Straten, A., van Oppen, P., Andersson, G., 2008. Are psychological and pharmacologic interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. *J. Clin. Psychiatry* 69, 1675–85; quiz 1839–41.
59. Cunningham, J.L., Wernroth, L., von Knorring, L., Berglund, L., Ekselius, L., 2011. Agreement between physicians' and patients' ratings on the Montgomery-Åsberg Depression Rating Scale. *J. Affect. Disord.* 135, 148–53.
60. Cusin, C., Yang, H., Yeung, A., Fava, M., 2009. Rating scales for depression, in: Baer, L., Blais, M.A. (Eds.), *Handbook of Clinical Rating Scales and Assessment in Psychiatry and Mental Health*. Humana Press, New York, NY, pp. 7–35.
61. Davies, B., Burrows, G., Poynton, C., 1975. A comparative study of four depression rating scales. *Aust. N. Z. J. Psychiatry* 9, 21–4.
62. De Bolle, M., De Fruyt, F., Quilty, L.C., Rolland, J.-P., Decuyper, M., Bagby, R.M., 2011. Does personality disorder co-morbidity impact treatment outcome for patients with major depression? A multi-level analysis. *J. Pers. Disord.* 25, 1–15.
63. DerSimonian, R., Laird, N., 1986. Meta-analysis in clinical trials. *Control. Clin. Trials* 7, 177–88.
64. Devillé, W.L., Buntinx, F., Bouter, L.M., Montori, V.M., de Vet, H.C.W., van der Windt, D.A.W.M., Bezemer, P.D., 2002. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med. Res. Methodol.* 2, 9.
65. Dittmann, V., 1991. Modern psychiatric classification in research and clinical practice. *Schweizer Arch. für Neurol. und Psychiatr.* (Zurich, Switz. 1985) 142, 341–53.
66. Doebler, P., 2015. *mada: Meta-Analysis of Diagnostic Accuracy*.
67. Doering, L. V, Cross, R., Magsarili, M.C., Howitt, L.Y., Cowan, M.J., 2007. Utility of observer-rated and self-report instruments for detecting major depression in women after cardiac surgery: a pilot study. *Am. J. Crit. Care* 16, 260–9.

68. Dratcu, L., da Costa Ribeiro, L., Calil, H.M., 1987. Depression assessment in Brazil. The first application of the Montgomery-Asberg Depression Rating Scale. *Br. J. Psychiatry* 150, 797–800.
69. Dunlop, B.W., Li, T., Kornstein, S.G., Friedman, E.S., Rothschild, A.J., Pedersen, R., Ninan, P., Keller, M., 2010. Correlation between patient and clinician assessments of depression severity in the PREVENT study. *Psychiatry Res.* 177, 177–183.
70. Duval, S., Tweedie, R., 2000a. A Nonparametric “ Trim and Fill ” Method of Accounting for Publication Bias in Meta-Analysis. *J. Am. Stat. Soc.* 95, 89–98.
71. Duval, S., Tweedie, R., 2000b. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56, 455–63.
72. Egger, M., Davey Smith, G., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315, 629–34.
73. El-Giamal, N., de Zwaan, M., Bailer, U., Strnad, A., Schüssler, P., Kasper, S., 2003. Milnacipran in the treatment of bulimia nervosa: a report of 16 cases. *Eur. Neuropsychopharmacol.* 13, 73–9.
74. Elizagarate, E., Cortes, J., Gonzalez Pinto, A., Gutierrez, M., Alonso, I., Alcorta, P., Ramirez, M., de Heredia, J.L., Figuerido, J.L., 2001. Study of the influence of electroconvulsive therapy on the regional cerebral blood flow by HMPAO-SPECT. *J. Affect. Disord.* 65, 55–9.
75. El-Khalili, N., Joyce, M., Atkinson, S., Buynak, R.J., Datto, C., Lindgren, P., Eriksson, H., 2010. Extended-release quetiapine fumarate (quetiapine XR) as adjunctive therapy in major depressive disorder (MDD) in patients with an inadequate response to ongoing antidepressant treatment: a multicentre, randomized, double-blind, placebo-controlled study. *Int. J. Neuropsychopharmacol.* 13, 917–32.
76. Enns, M.W., Larsen, D.K., Cox, B.J., 2000. Discrepancies between self and observer ratings of depression. The relationship to demographic, clinical and personality variables. *J. Affect. Disord.* 60, 33–41.
77. Erić-Marinković, J., Dotlić, R., Janošević, S., Kocev, N., Gajić, M., Ille, T., Stanisavljević, D., Babić, D., 2001. *Statistika za istraživače u oblasti medicinskih nauka*. Medicinski fakultet Univerziteta u Beogradu, Beograd.
78. Fajgelj, S., 2013. *Psihometrija: metod i teorija psihološkog merenja, IV dopunje*. ed. Centar za primenjenu psihologiju, Beograd.
79. Fava, M., McCall, W.V., Krystal, A., Wessel, T., Rubens, R., Caron, J., Amato, D., Roth, T., 2006. Eszopiclone co-administered with fluoxetine in patients with insomnia coexisting with major depressive disorder. *Biol. Psychiatry* 59, 1052–60.

80. Feliciano, L., Gum, A.M., 2009. Mood Disorders, in: Segal, D.L., Hersen, M. (Eds.), *Diagnostic Interviewing*. Springer, New York, Dordrecht, Heidelberg, London, pp. 153–176.
81. Fisher, R.A., 1921. On the probable error of a coefficient of correlation deduced from a small sample. *Metron* 1, 3–32.
82. Fisher, Z., Tipton, E., 2015. robumeta: Robust Variance Meta-regression. R package version 1.5. [WWW Document]. URL <http://cran.r-project.org/package=robumeta>
83. Fiske, A., O’Riley, A.A., 2008. Depression in Late Life, in: Hunsley, J., Mash, E.J. (Eds.), *A Guide to Assessments That Work*. Oxford University Press, USA.
84. Fleck, M.P., Poirier-Littre, M.F., Guelfi, J.D., Bourdel, M.C., Loo, H., 1995. Factorial structure of the 17-item Hamilton Depression Rating Scale. *Acta Psychiatr. Scand.* 92, 168–72.
85. Flint, A.J., Rifat, S.L., 1996. Validation of the hospital anxiety and depression scale as a measure of severity of geriatric depression. *Int. J. Geriatr. Psychiatry* 11, 991–994.
86. Galinowski, A., Lehert, P., 1995. Structural validity of MADRS during antidepressant treatment. *Int. Clin. Psychopharmacol.* 10, 157–61.
87. Gençöz, F., Gençöz, T., Soykan, A., 2007. Psychometric properties of the Hamilton Depression Rating Scale and other physician-rated psychiatric scales for the assessment of depression in ESRD patients undergoing hemodialysis in Turkey. *Psychol. Health Med.* 12, 450–9.
88. Glass, G. V, 1977. Integrating Findings: The Meta-Analysis of Research. *Rev. Res. Educ.* 5, 351–379.
89. Glass, G. V., 1976. Primary, Secondary, and Meta-Analysis of Research. *Educ. Res.* 5, 3–8.
90. Glasziou, P.P., Sanders, S.L., 2002. Investigating causes of heterogeneity in systematic reviews. *Stat. Med.* 21, 1503–11.
91. Gliner, J.A., Morgan, G.A., Leech, N.L., 2009. *Research Methods in Applied Settings: An Integrated Approach to Design and Analysis*. Routledge, New York, London.
92. González-Pinto, A., Mosquera, F., Reed, C., Novick, D., Barbeito, S., Vega, P., Bertsch, J., Alberich, S., Haro, J.M., 2009. Validity and reliability of the Hamilton depression rating scale (5 items) for manic and mixed bipolar disorders. *J. Nerv. Ment. Dis.* 197, 682–6.
93. Gordis, L., 2013. *Epidemiology*. Elsevier Health Sciences.
94. Greenhalgh, T., 1997. Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ* 315, 672–5.

95. Guilford, J.P., 1954. *Psychometric methods*. McGraw-Hill.
96. Guion, R.M., 1980. On Trinitarian doctrines of validity. *Prof. Psychol.*
97. Guttman, L., 1945. A basis for analyzing test-retest reliability. *Psychometrika* 10, 255–82.
98. Guyatt, G., 2014. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*, 3E. McGraw Hill Professional.
99. Hallgren, K.A., 2012. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor. Quant. Methods Psychol.* 8, 23–34.
100. Hamilton, M., 1967. Development of a rating scale for primary depressive illness. *Br. J. Soc. Clin. Psychol.* 6, 278–96.
101. Hamilton, M., 1960. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* 23, 56–62.
102. Hammen, C.L., 1980. Depression in college students: beyond the Beck Depression Inventory. *J. Consult. Clin. Psychol.* 48, 126–8.
103. Hedges, L.V., Olkin, I., 1982. Analyses, reanalyses, and meta-analysis. *Contemp. Educ. Rev.* 1, 157–165.
104. Hedges, L. V, 1981. Distribution theory for Glass's estimator of effect size and related estimators. *J. Educ. Stat.*
105. Hedges, L. V, Olkin, I., 1993. *Statistical Methods for Meta-analysis*. Academic Press, New York.
106. Hedges, L. V, Tipton, E., Johnson, M.C., 2010. Robust variance estimation in meta-regression with dependent effect size estimates. *Res. Synth. Methods* 1, 39–65.
107. Hedlund, J.L., Vieweg, B.W., 1979. The Hamilton Rating Scale for Depression: A comprehensive review. *J. Oper. Psychiatry* 10, 149–163.
108. Higgins, J.P.T., Green, S., 2008. *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons.
109. Higuchi, H., Sato, K., Yoshida, K., Takahashi, H., Kamata, M., Otani, K., Yamaguchi, N., 2008. No predictors of antidepressant patient response to milnacipran were obtained using the three-factor structures of the Montgomery and Asberg Depression Rating Scale in Japanese patients with major depressive disorders. *Psychiatry Clin. Neurosci.* 62, 197–202.
110. Hintikka, J., Tolmunen, T., Tanskanen, A., Viinamäki, H., 2003. High vitamin B12 level and good treatment outcome may be associated in major depressive disorder. *BMC Psychiatry* 3, 17.
111. Holzer, M., Padberg, F., 2010. Intermittent theta burst stimulation (iTBS) ameliorates therapy-resistant depression: a case series. *Brain Stimul.* 3, 181–3.

112. Hosang, G.M., Uher, R., Maughan, B., McGuffin, P., Farmer, A.E., 2012. The role of loss and danger events in symptom exacerbation in bipolar disorder. *J. Psychiatr. Res.* 46, 1584–9.
113. Huang, C.-C., Su, T.-P., Wei, I.-H., 2005. Repetitive transcranial magnetic stimulation for treating medication-resistant depression in Taiwan: a preliminary study. *J. Chin. Med. Assoc.* 68, 210–5.
114. Hunter, J.E., Schmidt, F.L., Jackson, G.B., 1982. *Meta-analysis: cumulating research findings across studies*, *Studying organizations: innovations in methodology*. Beverly Hills.
115. Iannuzzo, R.W., Jaeger, J., Goldberg, J.F., Kafantaris, V., Sublette, M.E., 2006. Development and reliability of the HAM-D/MADRS interview: an integrated depression symptom rating scale. *Psychiatry Res.* 145, 21–37.
116. Jarrett, R.B., Eaves, G.G., Grannemann, B.D., Rush, A.J., 1991a. Clinical, cognitive, and demographic predictors of response to cognitive therapy for depression: a preliminary report. *Psychiatry Res.* 37, 245–60.
117. Jarrett, R.B., Giles, D.E., Gullion, C.M., Rush, A.J., 1991b. Does learned resourcefulness predict response to cognitive therapy in depressed outpatients? *J. Affect. Disord.* 23, 223–9.
118. Jiang, Q., Ahmed, S., 2009. An analysis of correlations among four outcome scales employed in clinical trials of patients with major depressive disorder. *Ann. Gen. Psychiatry* 8, 4.
119. Johnson, B.T., Mullen, B., Salas, E., 1995. Comparison of three major meta-analytic approaches. *J. Appl. Psychol.* 80, 94–106.
120. Kang, H.-J., Stewart, R., Kim, J.-M., Jang, J.-E., Kim, S.-Y., Bae, K.-Y., Kim, S.-W., Shin, I.-S., Park, M.-S., Cho, K.-H., Yoon, J.-S., 2013. Comparative validity of depression assessment scales for screening poststroke depression. *J. Affect. Disord.* 147, 186–91.
121. Kaplan, R.M., Saccuzzo, D.P., 2013. *Psychological Testing: Principles, Applications, and Issues*. Wadsworth Publishing.
122. Keller, K.E., 1983. Dysfunctional attitudes and the cognitive therapy for depression. *Cognit. Ther. Res.* 7, 437–444.
123. Kennedy, R.S., Turnage, J.J., 1991. Reliability generalization: a viable key for establishing validity generalization. *Percept. Mot. Skills* 72, 297–8.
124. Khan, A., Brodhead, A.E., Kolts, R.L., 2004. Relative sensitivity of the Montgomery-Asberg depression rating scale, the Hamilton depression rating scale and the Clinical Global Impressions rating scale in antidepressant clinical trials: a replication analysis. *Int. Clin. Psychopharmacol.* 19, 157–60.

125. Kinsinger, S.W., Lattie, E., Mohr, D.C., 2010. Relationship between depression, fatigue, subjective cognitive impairment, and objective neuropsychological functioning in patients with multiple sclerosis. *Neuropsychology* 24, 573–80.
126. Kirch, W., 2008. *Encyclopedia of Public Health*. 2 Vol., Volume 1. Springer Science & Business Media.
127. Kobak, K.A., Leuchter, A., DeBrot, D., Engelhardt, N., Williams, J.B.W., Cook, I.A., Leon, A.C., Alpert, J., 2010. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. *J. Clin. Psychopharmacol.* 30, 193–7.
128. Koivumaa-Honkanen, H., Honkanen, R., Antikainen, R., Hintikka, J., Laukkanen, E., Honkalampi, K., Viinamäki, H., Viinamaeki, H., 2001. Self-reported life satisfaction and recovery from depression in a 1-year prospective study. *Acta Psychiatr. Scand.* 103, 38–44.
129. Kreuzer, P.M., Langguth, B., Schecklmann, M., Eichhammer, P., Hajak, G., Landgrebe, M., 2012. Can repetitive transcranial magnetic stimulation prolong the antidepressant effects of sleep deprivation? *Brain Stimul.* 5, 141–7.
130. Kuder, G.F., Richardson, M.W., 1937. The theory of the estimation of test reliability. *Psychometrika* 2, 151–160.
131. Kuhlman, I., Tolvanen, A., Seikkula, J., 2013. Couple Therapy for Depression Within a Naturalistic Setting in Finland: Factors Related to Change of the Patient and the Spouse. *Contemp. Fam. Ther. An Int. J.* 35, 656–672.
132. Laasonen-Balk, T., Viinamäki, H., Kuikka, J.T., Husso-Saastamoinen, M., Lehtonen, J., Tiihonen, J., 2004. 123I-beta-CIT binding and recovery from depression. A six-month follow-up study. *Eur. Arch. Psychiatry Clin. Neurosci.* 254, 152–5.
133. Laenen, A., Alonso, A., Molenberghs, G., Vangeneugden, T., Mallinckrodt, C.H., 2009. Using longitudinal data from a clinical trial in depression to assess the reliability of its outcome scales. *J. Psychiatr. Res.* 43, 730–8.
134. Lako, I.M., Bruggeman, R., Kneegting, H., Wiersma, D., Schoevers, R.A., Slooff, C.J., Taxis, K., 2012. A systematic review of instruments to measure depressive symptoms in patients with schizophrenia. *J. Affect. Disord.* 140, 38–47.
135. Laska, A.C., Mårtensson, B., Kahan, T., von Arbin, M., Murray, V., 2007. Recognition of depression in aphasic stroke patients. *Cerebrovasc. Dis.* 24, 74–9.
136. Lau, T.S., 1989. On repeated screening tests. *Biometrics* 45, 891–898.
137. Lecrubier, Y., Bech, P., 2007. The Ham D(6) is more homogenous and as sensitive as the Ham D(17). *Eur. Psychiatry* 22, 252–5.

138. Lempérière, T., Lépine, J.P., Rouillon, F., Hardy, P., Ades, J., Luauté, J.P., Ferrand, I., 1984. Comparaison de différents instruments d'évaluation de la dépression, à l'occasion d'une étude sur l'ATHYMIL 30 mg. *Ann. Médico-psychologiques, Rev. Psychiatr.* 142, 1206–14.
139. Leyman, L., De Raedt, R., Vanderhasselt, M.-A., Baeken, C., 2011. Effects of repetitive transcranial magnetic stimulation of the dorsolateral prefrontal cortex on the attentional processing of emotional information in major depression: a pilot study. *Psychiatry Res.* 185, 102–7.
140. Licht, R.W., Qvitzau, S., Allerup, P., Bech, P., 2005. Validation of the Bech-Rafaelsen Melancholia Scale and the Hamilton Depression Scale in patients with major depression; is the total score a valid measure of illness severity? *Acta Psychiatr. Scand.* 111, 144–9.
141. Lin, C.-H., Lu, M.-J., Wong, J., Chen, C.-C., 2014. Comparison of physician-rating and self-rating scales for patients with major depressive disorder. *J. Clin. Psychopharmacol.* 34, 716–21.
142. Lipsey, M., 2011. Meta-Analysis: What, Why, and the Basics of How [WWW Document]. URL <http://kc.vanderbilt.edu/quant/Seminar/LipseyHandoutJan11.pdf>
143. Littell, J., Corcoran, J., Pillai, V., 2008. *Systematic Reviews and Meta-Analysis*. Oxford University Press.
144. Loevinger, J., 1948. The technic of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychol. Bull.* 45, 507–29.
145. Lord, F.M., 1959. Statistical inferences about true scores. *Psychometrika* 24, 1–17.
146. Luborzewski, A., Schubert, F., Seifert, F., Danker-Hopfe, H., Brakemeier, E.-L., Schlattmann, P., Anghelescu, I., Colla, M., Bajbouj, M., 2007. Metabolic alterations in the dorsolateral prefrontal cortex after treatment with high-frequency repetitive transcranial magnetic stimulation in patients with unipolar major depression. *J. Psychiatr. Res.* 41, 606–15.
147. Machin, D., Campbell, M.J., 2005. *The Design of Studies for Medical Research*.
148. Maier, W., Philipp, M., Gerken, A., 1985. Dimensions of the Hamilton Depression Scale. Factor analysis studies. *Eur. Arch. Psychiatry Neurol. Sci.* 234, 417–22.
149. Maier, W., Philipp, M., Heuser, I., Schlegel, S., Buller, R., Wetzell, H., 1988. Improving depression severity assessment--I. Reliability, internal validity and sensitivity to change of three observer depression scales. *J. Psychiatr. Res.* 22, 3–12.

150. Malone, D.A., Dougherty, D.D., Rezai, A.R., Carpenter, L.L., Friehs, G.M., Eskandar, E.N., Rauch, S.L., Rasmussen, S.A., Machado, A.G., Kubu, C.S., Tyrka, A.R., Price, L.H., Stypulkowski, P.H., Giftakis, J.E., Rise, M.T., Malloy, P.F., Salloway, S.P., Greenberg, B.D., 2009. Deep brain stimulation of the ventral capsule/ventral striatum for treatment-resistant depression. *Biol. Psychiatry* 65, 267–75.
151. Malone, J.R., Hemsley, D.R., 1977. Lowered responsiveness and auditory signal detectability during depression. *Psychol. Med.* 7, 717–22.
152. Małyszczak, K., Inglot, M., Pawłowski, T., Czarnecki, M., Rymer, W., Andrzej, K., 2006. [Depressive symptoms during treatment with interferon alpha for HCV infection--preliminary report]. *Psychiatr. Pol.* 40, 799–808.
153. Martinsen, E.W., Friis, S., Hoffart, A., 1995. Assessment of depression: comparison between Beck Depression Inventory and subscales of Comprehensive Psychopathological Rating Scale. *Acta Psychiatr. Scand.* 92, 460–3.
154. Martiny, K., Lunde, M., Simonsen, C., Clemmensen, L., Poulsen, D.L., Solstad, K., Bech, P., 2004. Relapse prevention by citalopram in SAD patients responding to 1 week of light therapy. A placebo-controlled study. *Acta Psychiatr. Scand.* 109, 230–4.
155. Mbaya, P., 2002. Safety and efficacy of high dose of venlafaxine XL in treatment resistant major depression. *Hum. Psychopharmacol.* 17, 335–9.
156. McDowell, I., 2006. *Measuring Health: A Guide to Rating Scales and Questionnaires, Volume 1.* Oxford University Press, New York.
157. McIntyre, R.S., Konarski, J.Z., Mancini, D.A., Fulton, K.A., Parikh, S. V., Grigoriadis, S., Grupp, L.A., Bakish, D., Filteau, M.J., Gorman, C., Nemeroff, C.B., Kennedy, S.H., 2005. Measuring the severity of depression and remission in primary care: Validation of the HAM-D-7 scale. *Cmaj* 173, 1327–1331.
158. Meader, N., Moe-Byrne, T., Llewellyn, A., Mitchell, A.J., 2014. Screening for poststroke major depression: a meta-analysis of diagnostic validity studies. *J. Neurol. Neurosurg. Psychiatry* 85, 198–206.
159. Merkl, A., Schneider, G.-H., Schönecker, T., Aust, S., Kühl, K.-P., Kupsch, A., Kühn, A.A., Bajbouj, M., 2013. Antidepressant effects after short-term and chronic stimulation of the subgenual cingulate gyrus in treatment-resistant depression. *Exp. Neurol.* 249, 160–8.
160. Miles, J., Gilbert, P., 2005. *A Handbook of Research Methods for Clinical and Health Psychology.* Oxford University Press, Oxford.
161. Miller, I.W., Bishop, S.B., Norman, W.H., Keitner, G.I., 1985. Cognitive/Behavioural Therapy and Pharmacotherapy with Chronic, Drug-Refractory Depressed Inpatients: A Note of Optimism. *Behav. Psychother.* 13, 320.

162. Miller, L., Gur, M., Shanok, A., Weissman, M., 2008. Interpersonal psychotherapy with pregnant adolescents: two pilot studies. *J. Child Psychol. Psychiatry.* 49, 733–42.
163. Mimura, C., Murashige, M., Oda, T., Watanabe, Y., 2011. Development and psychometric evaluation of a Japanese scale to assess depression severity: Himorogi Self-rating Depression Scale. *Int. J. Psychiatry Clin. Pract.* 15, 50–5.
164. Mitchell, A.J., Bird, V., Rizzo, M., Meader, N., 2010. Which version of the geriatric depression scale is most useful in medical settings and nursing homes? Diagnostic validity meta-analysis. *Am. J. Geriatr. Psychiatry* 18, 1066–77.
165. Mitchell, M., Jolley, J., 2012. *Research Design Explained.* Cengage Learning, Belmont, CA.
166. Mittmann, N., Mitter, S., Borden, E.K., Herrmann, N., Naranjo, C.A., Shear, N.H., 1997. Montgomery-Asberg severity gradations. *Am. J. Psychiatry* 154, 1320–1.
167. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 339, b2535.
168. Mokken, R.J., 1971. *A Theory and Procedure of Scale Analysis: With Applications in Political Research.* Walter de Gruyter.
169. Möller, H.J., 2000. Rating depressed patients: observer- vs self-assessment. *Eur. Psychiatry* 15, 160–72.
170. Möller, H.-J., 2014. Observer rating scales, in: Alexopoulos, G., Kasper, S., Möller, H.-J., Moreno, C. (Eds.), *Guide to Assessment Scales in Major Depressive Disorder.* Springer, Heidelberg.
171. Möller, H.-J., 2009. Standardised rating scales in psychiatry: methodological basis, their possibilities and limitations and descriptions of important rating scales. *World J. Biol. Psychiatry* 10, 6–26.
172. Möller, H.-J., Schnitker, J., 2007. Prospective study using a modified Montgomery-Asberg Depression Scale. *Nervenarzt* 78, 685–90.
173. Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. *Br. J. Psychiatry* 134, 382–9.
174. Moras, K., di Nardo, P. a., Barlow, D.H., 1992. Distinguishing anxiety and depression: Reexamination of the reconstructed Hamilton scales. *Psychol. Assess.* 4, 224–227.
175. Morgado, A., Raoux, N., Smith, M., Allilaire, J.F., Widlöcher, D., 1989. Subjective bias in reports of poor work adjustment in depressed patients. *Acta Psychiatr. Scand.* 80, 541–7.

176. Morris, C.D., Miklowitz, D.J., Wisniewski, S.R., Giese, A.A., Thomas, M.R., Allen, M.H., 2005. Care satisfaction, hope, and life functioning among adults with bipolar disorder: data from the first 1000 participants in the Systematic Treatment Enhancement Program. *Compr. Psychiatry* 46, 98–104.
177. Morris, C.N., 2012. Parametric Empirical Bayes Inference: Theory and Applications. *J. Am. Stat. Assoc.*
178. Morriss, R., Leese, M., Chatwin, J., Baldwin, D., 2008. Inter-rater reliability of the Hamilton Depression Rating Scale as a diagnostic and outcome measure of depression in primary care. *J. Affect. Disord.* 111, 204–13.
179. Moses, L.E., Shapiro, D., Littenberg, B., 1993. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat. Med.* 12, 1293–316.
180. Moss, P.A., 1994. Can There Be Validity without Reliability? *Educ. Res.* 23, 5–12.
181. Muchinsky, P.M., 2006. *Psychology Applied to Work: An Introduction to Industrial and Organizational Psychology, Volume 1.* Cengage Learning, Belmont, CA.
182. Myers, K., Winters, N.C., 2002. Ten-year review of rating scales. I: overview of scale functioning, psychometric properties, and selection. *J. Am. Acad. Child Adolesc. Psychiatry* 41, 114–22.
183. Nezu, A.M., Nezu, C.M., Felgoise, S.H., McClure, K.S., Houts, P.S., 2003. Project Genesis: assessing the efficacy of problem-solving therapy for distressed adult cancer patients. *J. Consult. Clin. Psychol.* 71, 1036–48.
184. Novick, M.R., 1966. The axioms and principal results of classical test theory. *J. Math. Psychol.* 3, 1–18.
185. Nunnally, J., 1978. *Psychometric theory*, 2nd ed. McGraw Hill, New York.
186. Oxman, T.E., Hull, J.G., 1997. Social support, depression, and activities of daily living in older heart surgery patients. *J. Gerontol. B. Psychol. Sci. Soc. Sci.* 52B, P1–14.
187. Palmio, J., Huuhka, M., Laine, S., Huhtala, H., Peltola, J., Leinonen, E., Suhonen, J., Keränen, T., 2010. Electroconvulsive therapy and biomarkers of neuronal injury and plasticity: Serum levels of neuron-specific enolase and S-100b protein. *Psychiatry Res.* 177, 97–100.
188. Peyre, F., Martinez, R., Calache, M., Verdoux, H., Bourgeois, M., 1989. [New validation of the Montgomery and Asberg Depression Scale (MADRS) on a sample of 147 hospitalized depressed patients]. *Ann. médico-psychologiques* 147, 762–7.
189. Pintor, L., Bailles, E., Valldeoriola, F., Tolosa, E., Marti, M.J., de Pablo, J., 2006. Response to 4-month treatment with reboxetine in Parkinson's disease patients with a major depressive episode. *Gen. Hosp. Psychiatry* 28, 59–64.

190. Portney, L.G., Watkins, M.P., 2009. *Foundations of Clinical Research: Applications to Practice*, 3rd ed. Pearson Prentice Hall, Upper Saddle River, N.J.
191. Quintana, D.S., 2015. From pre-registration to publication: a non-technical primer for conducting a meta-analysis to synthesize correlational data. *Front. Psychol.* 6, 1549.
192. R Core Team, 2014. *R: A language and environment for statistical computing*.
193. Ramasubbu, R., Anderson, S., Haffenden, A., Chavda, S., Kiss, Z.H.T., 2013. Double-blind optimization of subcallosal cingulate deep brain stimulation for treatment-resistant depression: a pilot study. *J. Psychiatry Neurosci.* 38, 325–32.
194. Ramsey, F., Schafer, D., 2012. *The Statistical Sleuth: A Course in Methods of Data Analysis*. Cengage Learning.
195. Rocca, P., Fonzo, V., Ravizza, L., Rocca, G., Scotta, M., Zanaldi, E., Bogetto, F., 2002. A comparison of paroxetine and amisulpride in the treatment of dysthymic disorder. *J. Affect. Disord.* 70, 313–7.
196. Rosenberg, O., Shoenfeld, N., Zangen, a., Kotler, M., Dannon, P.N., 2010a. Deep TMS in a resistant major depressive disorder: A brief report. *Depress. Anxiety* 27, 465–469.
197. Rosenberg, O., Zangen, A., Stryker, R., Kotler, M., Dannon, P.N.P., 2010b. Response to deep TMS in depressive patients with previous electroconvulsive treatment. *Brain Stimul.* 3, 211–217.
198. Rosenthal, R., 1991. *Meta-analytic procedures for social research*. SAGE Publications, Inc.
199. Rosenthal, R., 1978. Combining results of independent studies. *Psychol. Bull.* 85, 185–193.
200. Rosenthal, R., Rubin, D.B., 1982a. Comparing effect sizes of independent studies. *Psychol. Bull.* 92, 500–504.
201. Rosenthal, R., Rubin, D.B., 1982b. Further meta-analytic procedures for assessing cognitive gender differences. *J. Educ. Psychol.* 74, 708–712.
202. Rothman, K., Greenland, S., Lash, T., 2008. *Modern Epidemiology*. Lippincott Williams & Wilkins.
203. Rothstein, H.R., Sutton, A.J., Borenstein, M., 2006. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. John Wiley & Sons.
204. Rotter, A., Biermann, T., Stark, C., Decker, A., Demling, J., Zimmermann, R., Sperling, W., Kornhuber, J., Henkel, A., 2013. Changes of Cytokine Profiles During Electroconvulsive Therapy in Patients With Major Depression. *J. ECT* 29, 162–169.

205. Safrit, M.J., Wood, T.M., 1995. Introduction to Measurement in Physical Education and Exercise Science. Mosby, St. Louis, MO.
206. Santor, D.A., Kusumakar, V., 2001. Open Trial of Interpersonal Therapy in Adolescents With Moderate to Severe Major Depression: Effectiveness of Novice IPT Therapists. *J. Am. Acad. Child Adolesc. Psychiatry* 40, 236–240.
207. Sarkar, D., 2008. Lattice: Multivariate Data Visualization with R. Springer-Verlag New York, New York.
208. Sayer, N.A., Sackeim, H.A., Moeller, J.R., Prudic, J., Devanand, D.P., Coleman, E.A., Kiersky, J.E., 1993. The relations between observer-rating and self-report of depressive symptomatology. *Psychol. Assess.* 5, 350–360.
209. Schennach, R., Obermeier, M., Seemüller, F., Jäger, M., Schmauss, M., Laux, G., Pfeiffer, H., Naber, D., Schmidt, L.G., Gaebel, W., Klosterkötter, J., Heuser, I., Maier, W., Lemke, M.R., Rütger, E., Klingberg, S., Gastpar, M., Riedel, M., Möller, H.-J., 2012. Evaluating depressive symptoms in schizophrenia: a psychometric comparison of the Calgary Depression Scale for Schizophrenia and the Hamilton Depression Rating Scale. *Psychopathology* 45, 276–85.
210. Schiffer, R.B., Wineman, N.M., 1990. Antidepressant pharmacotherapy of depression associated with multiple sclerosis. *Am. J. Psychiatry* 147, 1493–7.
211. Schlaepfer, T.E., Bewernick, B.H., Kayser, S., Mädler, B., Coenen, V.A., 2013. Rapid effects of deep brain stimulation for treatment-resistant major depression. *Biol. Psychiatry* 73, 1204–12.
212. Schmidt, F.L., Hunter, J.E., 2014. *Methods of Meta-Analysis: Correcting error and bias in research findings.* SAGE Publications.
213. Schneibel, R., Brakemeier, E.-L., Wilbertz, G., Dykier, P., Zobel, I., Schramm, E., 2012. Sensitivity to detect change and the correlation of clinical factors with the Hamilton Depression Rating Scale and the Beck Depression Inventory in depressed inpatients. *Psychiatry Res.* 198, 62–7.
214. Schrader, L.M., Cook, I.A., Miller, P.R., Maremont, E.R., DeGiorgio, C.M., 2011. Trigeminal nerve stimulation in major depressive disorder: first proof of concept in an open pilot trial. *Epilepsy Behav.* 22, 475–8.
215. Schrag, A., Barone, P., Brown, R.G., Leentjens, A.F.G., McDonald, W.M., Starkstein, S., Weintraub, D., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G.T., Goetz, C.G., 2007. Depression rating scales in Parkinson's disease: critique and recommendations. *Mov. Disord.* 22, 1077–92.
216. Schulze, R., 2007. Current Methods for Meta-Analysis. *Zeitschrift für Psychol. / J. Psychol.* 215, 90–103.
217. Schwarzer, G., 2014. meta: Meta-Analysis with R. R package version 3.8-0.

218. Schwarzer, G., Carpenter, J.R., Rücker, G., 2015. *Meta-Analysis with R*. Springer International Publishing.
219. Scogin, F., Beutler, L., Corbishley, A., Hamblin, D., 1988. Reliability and validity of the short form Beck Depression Inventory with older adults. *J. Clin. Psychol.* 44, 853–7.
220. Scott, C.S., Scott, J.L., Tacchi, M.J., Jones, R.H., 1994. Abbreviated Cognitive Therapy for Depression: A Pilot Study in Primary Care. *Behav. Cogn. Psychother.* 22, 57–64.
221. Seemüller, F., Riedel, M., Obermeier, M., Schennach-Wolff, R., Spellmann, I., Meyer, S., Bauer, M., Adli, M., Kronmüller, K., Ising, M., Brieger, P., Laux, G., Bender, W., Heuser, I., Zeiler, J., Gaebel, W., Möller, H.-J., 2012. The validity of self-rated psychotic symptoms in depressed inpatients. *Eur. Psychiatry* 27, 547–52.
222. Seikkula, J., Aaltonen, J., Kalla, O., Saarinen, P., Tolvanen, A., 2013. Couple therapy for depression in a naturalistic setting in Finland: a 2-year randomized trial. *J. Fam. Ther.* VO - 35 281.
223. Senra, C., 1996. Evaluation and monitoring of symptom severity and change in depressed outpatients. *J. Clin. Psychol.* 52, 317–324.
224. Senra, C., Polaino, A., 1993. Concordance between clinical and self-report depression scales during the acute phase and after treatment. *J. Affect. Disord.* 27, 13–9.
225. Shadish, W.R., Lecy, J.D., 2015. The meta-analytic big bang. *Res. Synth. Methods* 6, 246–64.
226. Shafer, A.B., 2006. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *J. Clin. Psychol.* 62, 123–46.
227. Shaffer, C.S., Shapiro, J., Sank, L.I., Coghlan, D.J., 1981. Positive changes in depression, anxiety, and assertion following individual and group cognitive behavior therapy intervention. *Cognit. Ther. Res.* 5, 149–157.
228. Sheehan, M.J., 1981. Constructs and “conflict” in depression. *Br. J. Psychol.* 72, 197–209.
229. Sidik, K., Jonkman, J.N., 2005. A note on variance estimation in random effects meta-regression. *J. Biopharm. Stat.* 15, 823–38.
230. Solyom, L., Solyom, C., Ledwidge, B., 1990. The fluoxetine treatment of low-weight, chronic bulimia nervosa. *J. Clin. Psychopharmacol.* 10, 421–5.
231. Spearman, C., 2010. The proof and measurement of association between two things. *Int. J. Epidemiol.* 39, 1137–50.
232. Spearman, C., 1910. Correlation calculated from faulty data. *Br. J. Psychol.* 1904-1920 3, 271–295.

233. Spielmans, G.I., Pasek, L.F., McFall, J.P., 2007. What are the active ingredients in cognitive and behavioral psychotherapy for anxious and depressed children? A meta-analytic review. *Clin. Psychol. Rev.* 27, 642–54.
234. Stage, K.B., Middelboe, T., Pisinger, C., 2003. Measurement of depression in patients with chronic obstructive pulmonary disease (COPD). *Nord. J. Psychiatry* 57, 297–301.
235. Stefanis, C.N., Stefanis, N.C., 2003. Diagnosis of Depressive Disorders: A Review, in: Maj, M., Sartorius, N. (Eds.), *Depressive Disorders*. John Wiley & Sons, pp. 1–51.
236. Stein, D.J., Lund, C., Nesse, R.M., 2013. Classification systems in psychiatry: diagnosis and global mental health in the era of DSM-5 and ICD-11. *Curr. Opin. Psychiatry* 26, 493–7.
237. Steinmeyer, E.M., Möller, H.J., 1992. Facet theoretic analysis of the Hamilton-D scale. *J. Affect. Disord.* 25, 53–61.
238. Sterne, J.A., Egger, M., 2001. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J. Clin. Epidemiol.* 54, 1046–55.
239. Stewart, L.A., Clarke, M.J., 1995. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Cochrane Working Group. Stat. Med.* 14, 2057–79.
240. Stockings, E., Degenhardt, L., Lee, Y.Y., Mihalopoulos, C., Liu, A., Hobbs, M., Patton, G., 2015. Symptom screening scales for detecting major depressive disorder in children and adolescents: a systematic review and meta-analysis of reliability, validity and diagnostic utility. *J. Affect. Disord.* 174, 447–63.
241. Stommel, M., Wills, C., 2004. *Clinical Research: Concepts and Principles for Advanced Practice Nurses*. Lippincott Williams & Wilkins.
242. Streiner, D.L., Norman, G.R., 2014. *Health Measurement Scales: A Practical Guide to Their Development and Use*, 5th ed. Oxford University Press, New York, NY.
243. Strober, M., DeAntonio, M., Schmidt-Lackner, S., Pataki, C., Freeman, R., Rigali, J., Rao, U., 1999. The pharmacotherapy of depressive illness in adolescents: an open-label comparison of fluoxetine with imipramine-treated historical controls. *J. Clin. Psychiatry* 60, 164–9.
244. Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A., Thacker, S.B., 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA* 283, 2008–12.

245. Tabuse, H., Kalali, A., Azuma, H., Ozaki, N., Iwata, N., Naitoh, H., Higuchi, T., Kanba, S., Shioe, K., Akechi, T., Furukawa, T.A., 2007. The new GRID Hamilton Rating Scale for Depression demonstrates excellent inter-rater reliability for inexperienced and experienced raters before and after training. *Psychiatry Res.* 153, 61–7.
246. Tanner-Smith, E.E., Tipton, E., 2014. Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and spss. *Res. Synth. Methods* 5, 13–30.
247. Thompson, C., 1989. Descriptive dimensions of rating scales, in: Thompson, C. (Ed.), *The Instruments of Psychiatric Research*. John Wiley & Sons, Chichester, pp. 5–17.
248. Thompson, W.M., Harris, B., Lazarus, J., Richards, C., 1998. A comparison of the performance of rating scales used in the diagnosis of postnatal depression. *Acta Psychiatr. Scand.* 98, 224–7.
249. Thunedborg, K., Black, C.H., Bech, P., 1995. Beyond the Hamilton depression scores in long-term treatment of manic-melancholic patients: prediction of recurrence of depression by quality of life measurements. *Psychother. Psychosom.* 64, 131–40.
250. Timotijević, I., Paunović, V.R., 2003. *Instrumenti kliničke procene u psihijatriji*. Institut za mentalno zdravlje, Beograd.
251. Tipton, E., 2014. Small sample adjustments for robust variance estimation with meta-regression. *Psychol. Methods*.
252. Toh, S., Hernán, M.A., 2008. Causal inference from longitudinal studies with baseline randomization. *Int. J. Biostat.* 4, Article 22.
253. Trajković, G., Starčević, V., Latas, M., Leštarević, M., Ille, T., Bukumirić, Z., Marinković, J., 2011. Reliability of the Hamilton Rating Scale for Depression: a meta-analysis over a period of 49 years. *Psychiatry Res.* 189, 1–9.
254. Tsai, A.C., 2014. Reliability and validity of depression assessment among persons with HIV in sub-Saharan Africa: systematic review and meta-analysis. *J. Acquir. Immune Defic. Syndr.* 66, 503–11.
255. Tsai, A.C., Scott, J.A., Hung, K.J., Zhu, J.Q., Matthews, L.T., Psaros, C., Tomlinson, M., 2013. Reliability and validity of instruments for assessing perinatal depression in African settings: systematic review and meta-analysis. *PLoS One* 8, e82521.
256. Türkçapar, M.H., Akdemir, A., Orsel, S.D., Demirergi, N., Sirin, A., Kiliç, E.Z., Ozbay, M.H., 1999. The validity of diagnosis of melancholic depression according to different diagnostic systems. *J. Affect. Disord.* 54, 101–7.

257. Uher, R., Farmer, A., Maier, W., Rietschel, M., Hauser, J., Marusic, A., Mors, O., Elkin, A., Williamson, R.J., Schmael, C., Henigsberg, N., Perez, J., Mendlewicz, J., Janzing, J.G.E., Zobel, A., Skibinska, M., Kozel, D., Stamp, A.S., Bajcs, M., Placentino, A., Barreto, M., McGuffin, P., Aitchison, K.J., 2008. Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychol. Med.* 38, 289–300.
258. Uher, R., Perlis, R.H., Placentino, A., Dernovšek, M.Z., Henigsberg, N., Mors, O., Maier, W., McGuffin, P., Farmer, A., 2012. Self-report and clinician-rated measures of depression severity: can one replace the other? *Depress. Anxiety* 29, 1043–9.
259. Urbina, M., Pineda, S., Piñango, L., Carreira, I., Lima, L., 1999. [3H]Paroxetine binding to human peripheral lymphocyte membranes of patients with major depression before and after treatment with fluoxetine. *Int. J. Immunopharmacol.* 21, 631–46.
260. Urbina, S., 2014. *Essentials of Psychological Testing*. John Wiley & Sons, Hoboken, NJ.
261. Vacha-Haase, T., 1998. Reliability Generalization: Exploring Variance in Measurement Error Affecting Score Reliability Across Studies. *Educ. Psychol. Meas.* 58, 6–20.
262. van de Rest, O., van der Zwaluw, N., Beekman, A.T.F., de Groot, L.C.P.G.M., Geleijnse, J.M., 2010. The reliability of three depression rating scales in a general population of Dutch older persons. *Int. J. Geriatr. Psychiatry* 25, 998–1005.
263. Van Laere, K., Nuttin, B., Gabriels, L., Dupont, P., Rasmussen, S., Greenberg, B.D., Cosyns, P., 2006. Metabolic imaging of anterior capsular stimulation in refractory obsessive-compulsive disorder: a key role for the subgenual anterior cingulate and ventral striatum. *J. Nucl. Med.* 47, 740–747.
264. Vassar, M., Bradley, G., 2012. A reliability generalization meta-analysis of coefficient alpha for the Reynolds Adolescent Depression Scale. *Clin. Child Psychol. Psychiatry* 17, 519–27.
265. Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 36, 1–48.
266. Viechtbauer, W., 2005. Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *J. Educ. Behav. Stat.* 30, 261–293.
267. Wagner, S., Helmreich, I., Lieb, K., Tadić, A., 2011. Standardized rater training for the Hamilton Depression Rating Scale (HAMD₁₇) and the Inventory of Depressive Symptoms (IDS(C30)). *Psychopathology* 44, 68–70.
268. Waller, M., Griffin, M., 1984. Group therapy for depressed elders. *Geriatr. Nurs.* 5, 309–11.

269. Walters, S.J., 2009. *Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation: A Practical Guide to Analysis and Interpretation*. John Wiley & Sons.
270. Wang, H.R., Woo, Y.S., Ahn, H.S., Ahn, I.M., Kim, H.J., Bahk, W.-M., 2015. The validity of the mood disorder questionnaire for screening bipolar disorder: a meta-analysis. *Depress. Anxiety* 32, 527–38.
271. Wang, M.-D., Geisser, S., 2003. Optimal dichotomization for repeated screening tests. *Stat. Probab. Lett.* 62, 61–70.
272. Wassertheil-Smoller, S., Smoller, J., 2015. *Biostatistics and Epidemiology: A Primer for Health and Biomedical Professionals*.
273. Webb, N.M., Shavelson, R.J., Haertel, E.H., 2006. Reliability coefficients and generalizability theory, in: Rao, C.R., Sinharay, S. (Eds.), *Handbook of Statistics*. North Holland, pp. 81–124.
274. Weiss, M., Gaston, L., Propst, A., Wisebord, S., Zicherman, V., 1997. The role of the alliance in the pharmacologic treatment of depression. *J. Clin. Psychiatry* 58, 196–204.
275. Weiss, R.D., Griffin, M.L., Mirin, S.M., 1989. Diagnosing major depression in cocaine abusers: the use of depression rating scales. *Psychiatry Res.* 28, 335–43.
276. Wells, K.C., Hersen, M., Bellack, A.S., Himmelhoch, J., 1979. Social skills training in unipolar nonpsychotic depression. *Am. J. Psychiatry* 136, 1331–2.
277. Whisman, M.A., Strosahl, K., Fruzzetti, A.E., Schmaling, K.B., Et Al, 1989. A structured interview version of the Hamilton Rating Scale for Depression: Reliability and validity. *Psychol. Assess.* 1, 238–241.
278. Wickham, H., 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
279. Williams, J.B., 2001. Standardizing the Hamilton Depression Rating Scale: past, present, and future. *Eur. Arch. Psychiatry Clin. Neurosci.* 251 Suppl , II6–12.
280. Williams, J.B.W., Kobak, K.A., Bech, P., Engelhardt, N., Evans, K., Lipsitz, J., Olin, J., Pearson, J., Kalali, A., 2008. The GRID-HAMD: standardization of the Hamilton Depression Rating Scale. *Int. Clin. Psychopharmacol.* 23, 120–9.
281. Williams, R.A., Pruitt, S.D., Doctor, J.N., Epping-Jordan, J.E., Wahlgren, D.R., Grant, I., Patterson, T.L., Webster, J.S., Slater, M.A., Atkinson, J.H., 1998. The contribution of job satisfaction to the transition from acute to chronic low back pain. *Arch. Phys. Med. Rehabil.* 79, 366–74.
282. Woodward, M., 2013. *Epidemiology: Study Design and Data Analysis, Third Edition*.
283. Zamora, J., Abaira, V., Muriel, A., Khan, K., Coomarasamy, A., 2006. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med. Res. Methodol.* 6, 31.

284. Zimmerman, M., Posternak, M.A., Chelminski, I., 2004. Derivation of a definition of remission on the Montgomery-Asberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. *J. Psychiatr. Res.* 38, 577–582.
285. Zou, K.H., 2001. Comparison of correlated receiver operating characteristic curves derived from repeated diagnostic test data. *Acad. Radiol.* 8, 225–33.

Prilozi

Prilog 1 – R komande korišćene u analizi podataka

RSkripte sa komandama za analizu Alfa koeficijenta, Loevingerovog koeficijenta, ICC i Kappa koeficijenta

```
# Podešavanje radnog direktorijuma i učitavanje neophodnih paketa

setwd("E:/R podaci"); require(xlsx); require(robumeta);
require(plotrix); library(ggplot2); require(meta); require(metafor);
require(lattice)

# Učitavanje baze podataka

podaci<-read.xlsx("Cronbach alfa baza podataka.xlsx",1);attach(podaci)

# NAZIVI I OBJAŠNJENJE VARIJABLI #
#sqrtvar - kvadratni koren iz recipročne varijanse zed transformisanih
koeficijenata, definiše veličine krugova na grafikonu
#log_time - logaritamski transformisano vreme procene
#esize - zed transformisana veličina efekta
#StudyNo - varijabla koja definiše studije, najčešće redni brojevi
studija
#var - inverzna varijansa
#TIPSTUD2 - tip studije: metodološka ili klinička
#DGGRUPA2 - primarna dijagnoza: depresija ili ostale dijagnoze
#Kriterijum - postojanje kriterijuma za uključenje ispitanika u
studiju koji je u vezi sa težinom bolesti
#Primarno_sekundarno, podaci saopšteni u studiji ili dobijeni
sekundarnom analizom iz IPD

#Linijski grafikon

size<-rescale(sqrtvar, c(0.5, 20))

ggplot(podaci, aes(x=log_time, y=esize, group=StudyNo)) +
geom_line(size = 0.3) + geom_point(size=size, color="blue", alpha=0.2
)+ geom_abline(intercept = 0.805, slope = 0.173, color="red", linetype
= 5, size = 1.5) + expand_limits(x = c(-0.2, 2.7), y = c(-0.1, 2))+
labs( x = "Logaritam vremena", y= "zed transformisana vrednost alfa
koeficijenata za HAMD") + geom_rug(col="blue",alpha=.1)

# Regresiona jednačina svih studija sa svim merenjima

robu(formula = esize ~ log_time, data = podaci, studynum = StudyNo,
var.eff.size = var, modelweights = c("CORR"), rho = 0.8, small = TRUE)

# Robumeta univarijantni regresioni modeli

time_c <- group.center(log_time, StudyNo); robu(formula = esize ~
time_c, podaci, studynum = StudyNo, var.eff.size = var, modelweights =
c("CORR"), rho = 0.8, small = TRUE)

robu(formula = esize ~ STAROST, podaci, studynum = StudyNo,
var.eff.size = var, modelweights = c("CORR"), rho = 0.8, small = TRUE)
```

```

# Robumeta multivarijantni regresioni model
robu(formula = esize ~ time_c+ HAMD_skor_c+ HAMD_sd_c, podaci,
studynum = StudyNo, var.eff.size = var, modelweights = c("CORR"), rho
= 0.8, small = TRUE)

# Objedinjavanje, izračunavanje tau2 i I2
robu(formula = esize ~ 1, data = podaci, studynum = StudyNo,
var.eff.size = var, modelweights = c("CORR"), rho = 0.8, small = TRUE)

# Analiza senzitivnosti
robumeta<- robu(formula = esize ~ 1, data = podaci, studynum =
StudyNo, var.eff.size = var, modelweights = c("CORR"), rho = 0.8,
small = TRUE)
sensitivity(robumeta)

# Forest plot svih studija sa svim merenjima
robumeta<- robu(formula = effect.size ~ 1, data = podaci, studynum =
Study, var.eff.size = var.eff.size, rho = 0.8, small = TRUE)
forest.rob(robumeta, es.lab = "Sequance", study.lab = "Study")

# Učitavanje baza podataka za analizu na prvoj i poslednjoj proceni
podaci<-read.xlsx("Cronbach alfa baza podataka - prva
procena.xlsx",1);attach(podaci)

# Meta-analiza na prvoj i poslednjoj proceni
res<-rma(esize, var, data= podaci, method="DL")
summary(res)

# Forest plot na prvoj i poslednjoj proceni
forest(res)

# Levkasti dijagram
funnel(trimfill(res), ylim=c(0,0.3),xlim=c(-0.1,1.75), cex=2, pch=16,
pch.fill=21, xlab = "zed transformisana vrednost alfa koeficijenata za
HAMD", cex.lab=1.1, ylab= "Standardna greška", main= "a) prva
procena",cex.main=1.1, cex.axis=1)

# Eggerov test
regtest(res)

# Baujat plot
baujat(res, ylim=c(0,0.2),xlim=c(0,3), xlab = "Kvadriran Pearsonov
rezidual", cex.lab=0.6, ylab= "Uticaj na ukupan rezultat", main= "a)
prva procena",cex.main=0.7, cex.axis=0.6)

# Uklanjanje podataka iz radnog prostora R programskog okruženja
detach(podaci)

```

RSkripte sa komandama za analizu korelacija

Podešavanje radnog direktorijuma i učitavanje neophodnih paketa

```
setwd("E:/R podaci"); library(xlsx); require(robumeta);  
require(plotrix); library(ggplot2); require(meta); require(metafor)
```

Učitavanje baze podataka

```
podaci<-read.xlsx("Korelacije HAMD-BDI baza  
podataka.xlsx",1);attach(podaci)
```

NAZIVI I OBJAŠNJENJE VARIJABLI

```
#sqrtvar je kvadratni koren iz n-3, definiše veličine krugova na  
grafikonu  
#log_time - logaritamski transformisano vreme procene  
#zcorr - zed transformisana vrednost koeficijenata korelacije  
#StudyNo - varijabla koja definiše studije, najčešće redni brojevi  
studija  
#var - inverzna varijansa
```

Linijski grafikon

```
size<-rescale(sqrtvar, c(0.5, 30))  
  
ggplot(podaci, aes(x=log_time, y=zcorr, group=StudyNo)) + geom_line()  
+ geom_point(size=size, color="blue", alpha=0.2 )+  
geom_abline(intercept = 0.63, slope = 0.23, color="red", linetype = 5,  
size = 2) + expand_limits(x = c(-0.3, 2.7), y = c(-1, 3))+ labs(title  
= "HAMD i BDI", x = "Logaritam vremena", y= "zed transformisana  
vrednost koeficijenata korelacije HAMD i BDI")+ theme(plot.title =  
element_text(size = rel(2), colour = "blue"))
```

Regresiona jednačina svih studija sa svim merenjima

```
robu(formula = zcorr ~ log_time, data = podaci, studynum = StudyNo,  
var.eff.size = var, modelweights = c("CORR"), rho = 0.8, small = TRUE)
```

Robumeta univarijantni regresioni modeli

```
time_c <- group.center(log_time, StudyNo); robu(formula = zcorr ~  
time_c, podaci, studynum = StudyNo, var.eff.size = var, modelweights =  
c("CORR"), rho = 0.8, small = TRUE)
```

```
robu(formula = zcorr ~ Primarno_sekundarno, podaci, studynum =  
StudyNo, var.eff.size = var, modelweights = c("CORR"), rho = 0.8,  
small = TRUE)
```

Robumeta multivarijantni regresioni modeli

```
robu(formula = zcorr ~ HAMD_skor_c+ BDI_skor_c+HAMD_sd_c+ BDI_sd_c,  
podaci, studynum = StudyNo, var.eff.size = var, modelweights =  
c("CORR"), rho = 0.8, small = TRUE)
```



```

# Objedinjavanje, izračunavanje tau2 i I2
robu(formula = zcorr ~ 1, data = podaci, studynum = StudyNo,
var.eff.size = var, modelweights = c("CORR"), rho = 0.8, small = TRUE)

# Analiza senzitivnosti
robumeta<- robu(formula = zcorr ~ 1, data = podaci, studynum =
StudyNo, var.eff.size = var, modelweights = c("CORR"), rho = 0.8,
small = TRUE)
sensitivity(robumeta)

# Forest plot svih studija sa svim merenjima
robumeta<- robu(formula = effect.size ~ 1, data = podaci, studynum =
Study, var.eff.size = var.eff.size, rho = 0.8, small = TRUE)
forest.robu(robumeta, es.lab = "Sequence", study.lab = "Study")

# Učitavanje baza podataka za analizu na prvoj i poslednjoj proceni
podaci<-read.xlsx("Korelacije HAMD-BDI prvo merenje.xlsx", 1);
attach(podaci)

# Meta-analiza na prvoj i poslednjoj proceni
res<-rma(zcorr, var, data= podaci, method="DL")
summary(res)

# Forest plot na prvoj i poslednjoj proceni
forest(res)

# Levkasti dijagram korelacija na prvoj i poslednjoj proceni
funnel(trimfill(res), ylim=c(0,1),xlim=c(-1.5,3.5), cex=2, pch=16,
pch.fill=21, xlab = "zed transformisana vrednost koeficijenata
korelacije HAMD i BDI", cex.lab=1.1, ylab= "Standardna greška", main=
"a) prva procena",cex.main=1.1, cex.axis=1)

# Eggerov test
regtest(res)

# Baujat plot
baujat(res, ylim=c(0,0.4),xlim=c(0,17), xlab = "Kvadriran Pearsonov
rezidual", cex.lab=0.6, ylab= "Uticaj na ukupan rezultat", main= "a)
prva procena",cex.main=0.7, cex.axis=0.6)

```

RSkripte sa komandama za analizu dijagnostičke valjanosti

```
# Podešavanje radnog direktorijuma i učitavanje neophodnih paketa
```

```
setwd("E:/R podaci"); require(xlsx); require(mada)
```

```
# Učitavanje baze podataka
```

```
podaci<-read.xlsx("Dg HAMD prva procena.xlsx", 1); attach(podaci)
```

```
# Deskriptivna statistika
```

```
Mada <- madad(podaci)
```

```
print(Mada, digits = 3)
```

```
summary(Mada)
```

```
# Forest plot na prvoj i poslednjoj proceni
```

```
forest(Mada)
```

```
forest(Mada, type = "sens", xlab = "Sensitivity")
```

```
forest(Mada, type = "spec", xlab = "Specificity")
```

```
forest(Mada, type = "DOR", xlab = "DOR")
```

```
# SROC grafikon na prvoj i poslednjoj proceni
```

```
crosshair(podaci)
```

```
ROCellipse(podaci)
```

```
mSLROC(podaci, add = TRUE)
```

```
# Meta-analiza dijagnostičke tačnosti
```

```
MADAUNI<-(madauni(podaci, type = "DOR", method = "DSL", suppress = TRUE))
```

```
summary(MADAUNI)
```

```
forest(MADAUNI)
```

Prilog 2 – Studije u kojima su saopštene vrednosti alfa koeficijenta

Alfa koeficijenti za HAMD skalu

Studija	Vreme ocenjivanja	Veličina uzorka	Alfa koeficijent
Addington i sar., 1990	Time 1	50	0,660
	Time 2	41	0,780
Addington i sar., 1996	Time 1	112	0,730
	Time 2	89	0,790
Baumann, 1976	Tag 0	197	0,730
	Tag 10	194	0,860
	Tag 20	163	0,870
Bech i sar., 1997	Baseline	359	0,480
Psihijatrija	After 4 weeks	322	0,850
Bech i sar., 1997	Baseline	97	0,590
Opšta praksa	After 4 weeks	96	0,720
Bent-Hansen i sar., 2003	Baseline	230	0,420
	Week 3	204	0,630
	Week 6	190	0,810
	Week 8	148	0,670
	Week 12	139	0,690
Dunlop i sar., 2010	Baseline	1047	0,180
	Week 10	789	0,760
	Month 6	435	0,720
Gençöz i sar., 2007	Baseline	45	0,830
	Follow-up	45	0,810
González-Pinto i sar., 2009	Baseline	173	0,670
	4 week	173	0,800
Hintikka i sar., 2003	Baseline	115	0,740
	Follow-up	115	0,810
Kobak et al., 2010	Baseline	66	0,539
	Visit 3	62	0,677
	Endpoint	51	0,806

Studija	Vreme ocenjivanja	Veličina uzorka	Alfa koeficijent
Kuhlman i sar., 2013	Baseline	51	0,320
	6-month post-baseline	51	0,800
Laasonen-Balk i sar., 2004	Baseline	18	0,760
	Six-month follow-up	18	0,780
Licht i sar., 2005	Baseline	1629	0,140
	Week 2	1581	0,670
	Week 6	1459	0,830
Maier i sar., 1985	P1	107	0,720
	P3	88	0,830
McIntyre i sar., 2005	Visit 2	410	0,570
	Visit 3	374	0,776
	Visit 4	357	0,819
	Visit 5	337	0,840
	Visit 6	322	0,848
Mimura i sar., 2011	Baseline	204	0,898
	Second data collection	191	0,899
	Third data collection	178	0,896
	Fourth data collection	166	0,898
Moras i sar., 1992	Interview 1	267	0,800
	Interview 2	267	0,830
Oxman i Hull, 1997	Presurgery	147	0,750
	1-month postsurgery	147	0,800
	6-months postsurgery	147	0,820
Sayer i sar., 1993	Pre-electroconvulsive therapy	114	0,720
	Post-sixth electroconvulsive therapy	99	0,870
	Post-electroconvulsive therapy	99	0,920

Studija	Vreme ocenjivanja	Veličina uzorka	Alfa koeficijent
Schennach i sar., 2012	Admission	278	0,720
	Discharge	278	0,790
Seikkula i sar., 2013	Outset	66	0,590
	12 month	51	0,866
Thunedborg i sar., 1995	First visit	23	0,820
	Second visit	23	0,850
Williams i sar., 1998	Baseline	82	0,840
	6 months	82	0,780

Alfa koeficijenti za MADRS skalu

Studija	Vreme ocenjivanja	Veličina uzorka	Alfa koeficijent
Bent-Hansen i sar., 2003	Baseline	230	-0,120
	Week 3	204	0,540
	Week 6	190	0,850
	Week 8	148	0,730
	Week 12	139	0,710
Bondolfi i sar., 2010	Time 1 (Study inclusion)	63	0,840
	Time 2 (4 weeks later)	63	0,910
Cunningham i sar., 2011	Week 0	390	0,590
	Week 2	366	0,800
	Week 4	346	0,870
	Week 8	338	0,890
	Week 12	326	0,890
	Week 16	316	0,900
	Week 20	312	0,900
	Week 24	334	0,910
De Bolle i sar., 2011	Pretreatment (Baseline)	562	0,690
	Posttreatment (After six months)	562	0,910
Galinowski i Lehert, 1995	Day 0	137	0,747
	Day 28	137	0,942
Morris i sar., 2005	Baseline	1000	0,890
	3-mo	1000	0,870
	6-mo	1000	0,860
	9-mo	1000	0,870
	12-mo	1000	0,870

Studija	Vreme ocenjivanja	Veličina uzorka	Alfa koeficijent
van de Rest i sar., 2010	Baseline	283	0,640
	After 13 week	283	0,720
	After 26 week	283	0,670

Prilog 3 – Studije u kojima su saopštene vrednosti Loevingerinog
koeficijenta

Loevingerini koeficijenti za HAMD skalu

Studija	Vreme ocenjivanja	Veličina uzorka	Loevingerin koeficijent
Bech i sar., 1997 (Psihijatrija)	Baseline	359	0,030
	After 4 weeks of therapy	322	0,300
Bech i sar., 1997 (Opšta praksa)	Baseline	97	0,070
	After 4 weeks of therapy	96	0,170
Bech i sar., 2004	After 4 week	419	0,240
	After 8 week	398	0,540
Bech i sar., 2009 (studija Fava i sar., 2006)	Week 4	481	0,340
	Week 8	458	0,350
Bech i sar., 2009 (studija Martiny i sar., 2004)	Week 1	85	0,300
	Week 3	82	0,390
	Week 9	76	0,390
Bent-Hansen i sar., 2003	Baseline	230	0,080
	Week 3	204	0,170
	Week 6	190	0,360
	Week 8	148	0,240
	Week 12	139	0,250
Lecrubier i Bech, 2007 (Hypericum)	After 4 week	186	0,240
	After 6 week	186	0,270
Lecrubier i Bech, 2007 (Placebo)	After 4 week	189	0,260
	After 6 week	189	0,290
Licht i sar., 2005	Baseline	1629	0,010
	2 weeks	1581	0,130
	4 weeks	1459	0,280
Thunedborg i sar., 1995	First visit	23	0,350
	Second visit	23	0,480

Loevingerini koeficijent za MADRS skalu

Studija	Vreme ocenjivanja	Veličina uzorka	Loevingerin koeficijent
Bech i sar., 2004	After 4 week	419	0,430
	After 8 week	398	0,540
Bent-Hansen i sar., 2003	Baseline	230	-0,010
	Week 3	204	0,130
	Week 6	190	0,420
	Week 8	148	0,270
	Week 12	139	0,260

Prilog 4 – Studije u kojima su saopštene ocene međuposmatračke
pouzdanosti ukupnih skorova

Ocene međuposmatračke pouzdanosti ukupnih skorova za HAMD skalu

Studija	Vreme procenjivanja	Broj ispitanika u studiji	Veličina uzorka na kojem je ispitana međuposmatračka pouzdanost	Način na koji je sprovedena procena intervjua ispitanika	Način izračunavanja međuposmatračke pouzdanosti	Koeficijent međuposmatračke pouzdanosti
Arian i sar., 1993	Pre tretmana	75				
	Posle tretmana					0,870
	U periodu			Audio zapis		0,840
	praćenja					0,840
Brakemeier i sar., 2011	Pre ECT	90			Intrarater reliability coefficients	
	Tokom tretmana					>0,980
	Posle ECT					
Castonguay i sar., 2004	Pretreatment	28	Approximately 25% of study participants	Audio recorded	ICC	0,740
	Post treatment	21				0,980
Cicchetti i Prusoff, 1983	At randomization	86	81	Independent interview	ICC	0,460
		81	79			0,820
	At end point					

Studija	Vreme procenjivanja	Broj ispitanika u studiji	Veličina uzorka na kojem je ispitana međuposmatračka pouzdanost	Način na koji je sprovedena procena intervjua ispitanika	Način izračunavanja međuposmatračke pouzdanosti	Koeficijent međuposmatračke pouzdanosti
Jarrett i sar., 1991b	Pretreatment	37	20	Video recorded	ICC	0,820
	Posttreatment		18			0,940
Kinsinger i sar., 2010	Baseline	127			ICC	0,890
	Posttreatment	122				(opseg, 0,75-0,97)
Kobak et al., 2010	Baseline	81	66		ICC	0,340
	Visit 3	51	62			0,480
	Endpoint		51			0,710
Laenen i sar., 2009 Trial 1	0	354			RT	0,350
	1		0,370			
	2		0,390			
	4		0,420			
	6		0,560			
	8		0,590			
Laenen i sar., 2009 Trial 2	10week	353			RT	0,660
	0		0,390			
	1		0,440			
	2		0,470			
	4		0,600			
	6		0,800			
8	0,750					
	10week				0,660	

Studija	Vreme procenjivanja	Broj ispitanika u studiji	Veličina uzorka na kojem je ispitana međuposmatračka pouzdanost	Način na koji je sprovedena procena intervjua ispitanika	Način izračunavanja međuposmatračke pouzdanosti	Koeficijent međuposmatračke pouzdanosti
Maier i sar., 1988	Pretreatment	178	48	Independent interview	ICC	0,720
	Posttreatment		48			0,700
Montgomery i Asberg, 1979	Before treatment	106	25	Joint interview		0,890
	During treatment		14			0,980
Morriss i sar., 2008	Session 1	42	42	Audio zapis	ICC	0,968
	Session 2					0,937
	Session 3					0,896
	Session 4					0,938
Nezu i sar., 2003	Baseline	132	46	Audio zapis		
	Posttreatment					0,890
	6-month follow-up					0,830
	1-year follow-up					0,880

Studija	Vreme procenjivanja	Broj ispitanika u studiji	Veličina uzorka na kojem je ispitana međuposmatračka pouzdanost	Način na koji je sprovedena procena intervjua ispitanika	Način izračunavanja međuposmatračke pouzdanosti	Koeficijent međuposmatračke pouzdanosti
Peyre i sar., 1989	Admission	147		Joint	Kappa	0,780
	Sortie	147		interview	Cohen	0,830
Santor i Kusumakar, 2001	Baseline	25	12	Joint	ICC	>0,850
	12-Week Follow-up		25	interview		>0,790
Strober i sar., 1999	At baseline	80	20	Joint	ICC	0,840
	Fluoxetine group	At week 6	20	interview		0,870
Strober i sar., 1999	At baseline	80	15	Joint	ICC	0,890
	Imipramine group	At week 6	15	interview		0,810
Tabuse i sar., 2007 Group A	Before training	3	3	Video recorded	ICC	0,950
	After training		3			0,950
Tabuse i sar., 2007 Group B	Before training	3	3	Video recorded	ICC	0,930
	After training		3			0,950
Tabuse i sar., 2007 Group C	Before training	3	3	Video recorded	ICC	0,970
	After training		3			0,950
Tabuse i sar., 2007 Group D	Before training	3		Video recorded	ICC	0,970
	After training		3			0,990

Studija	Vreme procenjivanja	Broj ispitanika u studiji	Veličina uzorka na kojem je ispitana međuposmatračka pouzdanost	Način na koji je sprovedena procena intervjua ispitanika	Način izračunavanja međuposmatračke pouzdanosti	Koeficijent međuposmatračke pouzdanosti
Wagner i sar., 2011	Rating 1	5	1	Video recorded	ICC	0,995
	Rating 5		1			0,999
Waller i Griffin, 1984	two three month segments	70	25	Joint interview	interrater reliability correlation	0,787
						0,998
Williams i sar., 2008	GRID-baseline	34	34	Independent interview	ICC	0,750
	GRID-week 4		31			0,810

Ocene međuposmatračke pouzdanosti ukupnih skorova za MADRS skalu

Studija	Vreme procenjivanja	Broj ispitanika u studiji	Veličina uzorka na kojem je ispitana međuposmatračka pouzdanost	Način na koji je sprovedena procena intervjua ispitanika	Način izračunavanja međuposmatračke pouzdanosti	Koeficijent međuposmatračke pouzdanosti
Bauer i sar., 2009	Baseline	493				0,845
	Follow-up	424			Kappa	0,850
Corruble i sar., 1998	D0	61	61	Video recorded		0,300
	D10	58	58		ICC	0,730
	D28	58	58			0,640
El-Khalili i sar., 2010	Baseline	446			Kappa	0,842
	Follow-up	259				0,739
Laenen i sar., 2009 Klinički ogled 1	0					0,360
	1					0,330
	2					0,390
	4	354				
	6				R _T	0,410
	8					0,500
						0,620
10week					0,630	

Studija	Vreme procenjivanja	Broj ispitanika u studiji	Veličina uzorka na kojem je ispitana međuposmatračka pouzdanost	Način na koji je sprovedena procena intervjua ispitanika	Način izračunavanja međuposmatračke pouzdanosti	Koeficijent međuposmatračke pouzdanosti
Laenen i sar., 2009	0					0,540
Klinički ogled 2	1					0,500
	2					0,490
	4	353			R _T	0,650
	6					0,820
	8					0,780
	10week					0,800
	Maier i sar., 1988	Pretreatment	178	48	Independent interview	ICC
	Post-treatment	48		0,820		
Montgomery i Asberg, 1979	Before treatment	106	25	Joint interview		0,890
	During treatment		14			0,950
Peyre i sar., 1989	Admission	147		Joint interview	Kappa	0,810
	Sortie	147				Cohen

Prilog 5 – Studije iz kojih su, sekundarnom analizom, izračunate ocene
test-retest pouzdanosti u ponovljenim merenjima

Sekundarnom analizom izračunate ocene test-retest pouzdanosti u ponovljenim merenjima za HAMD skalu

Studija	Vreme ocenjivanja	Veličina uzorka	Vreme između dva ocenjivanja (dani)	Način izračunavanja pouzdanosti	Test-retest koeficijent pouzdanosti
Kreuzer i sar., 2012	Day 0	28	1	<i>Pearson r</i>	0,441
	Day 1				
	Day 2	24			0,711
	Day 3				
Pintor i sar., 2006	T0	17	15	<i>Pearson r</i>	0,377
	T1 (15 days)				
	T2 (1 month)	15			0,670
	T3 (2 months)				

Prilog 6 – Studije u kojima su saopštene korelacije između skala HAMD,
MADRS i BDI

Korelacije između skala HAMD i BDI

Studija	Vreme ocenjivanja	Poreklo koeficijenata korelacije	Veličina uzorka	Koeficijent korelacije
Baeken i sar., 2009	Pre	IPD	21	0,670
	Post		21	0,754
Bailey i Coppen, 1976	On admission	Publikovani	42	0,330
	1st week		42	0,710
	2nd week		42	0,510
	3rd week		42	0,770
	4th week		24	0,790
Barefoot i sar., 2003	Baseline	Publikovani	196	0,730
	Follow-up (two weeks)		196	0,770
Bowen, 1978	Before	Publikovani	40	0,520
	After		40	0,570
Brown i sar., 1995	Month 1	Publikovani	166	0,700
	Month 2		151	0,770
	Month 3		142	0,810
	Month 4		119	0,750
	Month 6		104	0,840
	Month 8		103	0,850
Brunoni i sar., 2011	Baseline	IPD	5	-0,219
	Final		5	-0,179
Burrows i sar., 1972	Baseline	IPD	32	0,576
	Week 2		32	0,665
	Week 4		32	0,724
	Week 6		19	0,548
Chabrol i sar., 2002				0,910
	Baseline – 4-12 weeks		48	
	Outcome – 10-12 weeks postpartum	Publikovani	48	0,890

Studija	Vreme ocenjivanja	Poreklo koeficijenata korelacije	Veličina uzorka	Koeficijent korelacije
Collet i sar., 1987	Baseline		5	0,302
	After counter demands		5	0,660
	3 mo.	IPD	5	0,823
	6 mo.		5	0,931
	12 mo.		5	0,949
Corn i sar., 1984	Before surgery	IPD	6	-0,207
	After surgery		6	0,770
Davies et al., 1975	Day 0	Publikovani	72	0,730
	Day 7		24	0,660
	Day 14		18	0,550
	Day 21		10	0,750
El-Giamal i sar., 2003	Pre	IPD	16	0,227
	Post		16	0,183
Hammen, 1980	Baseline	Publikovani	27	0,800
	Follow-up		21	0,530
Holzer i Padberg, 2010	Baseline	IPD	7	0,618
	Week 1		7	0,699
	Week 2		7	0,736
	Week 3		7	0,827
Huang i sar., 2005	Baseline	IPD	11	0,466
	Week 1		10	0,781
	Week 2		10	0,759
Jarrett i sar., 1991a	Pretreatment	Publikovani	37	0,299
	Posttreatment		37	0,882
Keller, 1983	Pretreatment	Publikovani	18	0,420
	Posttreatment		18	0,820
	Follow-up		18	0,730

Studija	Vreme ocenjivanja	Poreklo koeficijenata korelacije	Veličina uzorka	Koeficijent korelacije
Koivumaa-Honkanen i sar., 2001	Baseline		185	0,599
	Month 6	Publikovani	166	0,770
	Month 12		157	0,872
Lempérière i sar., 1984	Day 0		75	0,590
	Day 7	Publikovani	73	0,700
	Day 14		74	0,800
	Day 28		74	0,830
Leyman i sar., 2011	Pre-rTMS	IPD	14	0,252
	Post-rTMS		14	0,697
Luborzewski i sar., 2007	Pre-rTMS	IPD	17	0,272
	Post-rTMS		17	0,585
Malone i Hemsley, 1977	During depression	Publikovani	10	0,313
	On remission		10	0,640
Małyszczak i sar., 2006	Before treatment	IPD	6	0,983
	After treatment		6	0,551
Merkl i sar., 2013	Baseline	IPD	6	0,412
	Month 6		6	0,956
Miller i sar., 1985	Pre-treatment	IPD	6	-0,794
	Post-treatment		6	0,982
Miller i sar., 2008	Week 1		11	0,499
	Week 12	IPD	11	0,749
	Week 20		11	0,819
Rosenberg i sar., 2010a				0,117
	Before study		7	0,864
	Session 5		7	0,951
	Session 10	IPD	7	0,987
	Session 15		6	0,518
	Session 20		5	

Studija	Vreme ocenjivanja	Poreklo koeficijenata korelacije	Veličina uzorka	Koeficijent korelacije
Rosenberg i sar., 2010b	Before study	IPD	6	0,388
	After 5 treatments		6	0,605
	After 10 treatments		6	0,727
	After 15 treatments		4	0,947
Sayer i sar., 1993	Pre-electroconvulsive therapy	Publikovani	114	0,482
	Post-sixth electroconvulsive therapy		99	0,766
	Post-electroconvulsive therapy		99	0,783
Schiffer i Wineman, 1990	Week 0	IPD	28	0,437
	Week 5		28	0,719
Schneibel i sar., 2012	Pre treatment	Publikovani	105	0,300
	Post treatment		105	0,810
Schrader i sar., 2011	Baseline	IPD	5	0,431
	Week 8		5	0,921
Scogin i sar., 1988	Pretreatment	Publikovani	54	0,420
	Posttreatment		54	0,540
Scott i sar., 1994	Pre-intervention	IPD	7	0,323
	Post-intervention		7	0,519
Senra i Polaino, 1993	Week 0	Publikovani	52	0,510
	Week 12		49	0,760
	Week 24		45	0,890
Shaffer i sar., 1981	Pre conditions	Publikovani	23	0,440
	Post conditions		23	0,540

Studija	Vreme ocenjivanja	Poreklo koeficijenata korelacije	Veličina uzorka	Koeficijent korelacije
Sheehan, 1981	Before treatment	Publikovani	13	0,710
	After treatment		13	0,680
Solyom i sar., 1990	Baseline	IPD	10	0,918
	At 3 months		10	0,730
Uher i sar., 2012	Baseline	Publikovani	811	0,480
	Exit		811	0,840
Urbina i sar., 1999	Before treatment	Publikovani	20	0,970
	After treatment		20	0,854
Van Laere i sar., 2006	Preoperative scores	IPD	6	0,744
	After chronic electrostimulation		6	0,970
Weiss i sar., 1997	At screening	Publikovani	31	0,430
	After treatment		31	0,750
Wells i sar., 1979	Pre-treatment	IPD	4	0,405
	Mild-treatment		4	0,874
	Post-treatment		4	0,971
Whisman i sar., 1989	Pretest	Publikovani	70	0,270
	Posttest		70	0,670

IPD: Individualni podaci pacijenta ; rTMS: Repetitivna transkranijalna magnetna stimulacija.

Korelacije između skala HAMD i MADRS

Studija	Vreme ocenjivanja	Poreklo koeficijenata korelacije	Veličina uzorka	Koeficijent korelacije
Bergman i sar., 2011	Baseline		13	0,954
	3 Week	IPD	13	0,903
	6 Week		13	0,958
Cook i sar., 1986	1st	IPD	15	0,942
	Final		15	0,939
Dratcu i sar., 1987	within 1 week from beginning of treatment	Publikovani	40	0,890
	after 4 week or at the end of an ECT series		10	0,996
Elizagarate i sar., 2001	Before ECT	IPD	10	0,238
	ECT		10	0,907
Flint i Rifat, 1996	Initial assessment	Publikovani	99	0,750
	Final assessment		99	0,890
Khan i sar., 2004	Baseline	Publikovani	121	0,560
	Final LOCF		116	0,960
Lempérière i sar., 1984	J0	Publikovani	79	0,690
	J7		79	0,860
	J14		79	0,920
	J28		79	0,920
Luborzewski i sar., 2007	Pre-rTMS	IPD	17	0,092
	Post-rTMS		17	0,523

Studija	Vreme ocenjivanja	Poreklo koeficijenata korelacije	Veličina uzorka	Koeficijent korelacije
Malone i sar., 2009	Baseline		15	0,685
	Month 1		15	0,867
	Month 3	IPD	15	0,917
	Month 6		15	0,873
	Month 12		11	0,885
Małyszczak i sar., 2006	Before treatment	IPD	6	0,972
	After treatment		6	0,980
Mbaya, 2002	Week 1		5	0,897
	Week 4		5	0,900
	Week 8		3	0,996
	Week 12	IPD	4	0,922
	Week 16		3	0,756
	Week 20		4	0,956
	Week 24		3	0,924
McIntyre i sar., 2005	Baseline	Publikovani	205	0,672
	Endpoint		157	0,923
Merkl i sar., 2013	Baseline	Publikovani	6	0,112
	Month 6		6	0,938

Studija	Vreme ocenjivanja	Poreklo koeficijenata korelacije	Veličina uzorka	Koeficijent korelacije
Mittmann i sar., 1997	Day 0		77	0,660
	Day 7	Publikovani	-	0,830
	Day 14		-	0,900
	Day 28		49	0,930
Morgado i sar., 1989	pretreatment	Publikovani	25	0,500
	posttreatment		25	0,770
Ramasubbu i sar., 2013	Pre-DBS		4	0,924
	Week 14		4	0,876
	Week 16		4	0,723
	Week 18		4	0,097
	Week 20		4	0,559
	Week 22		4	0,893
	Week 24	IPD	4	0,734
	Week 26		4	0,925
	Week 28		4	0,904
	Week 30		4	0,788
	Week 32		4	0,958
	Week 34		4	0,874
	Week 36		4	0,890
	Senra, 1996	Admission		52
12 week		Publikovani	49	0,830
24 week			45	0,880
Uher i sar., 2012	Baseline	Publikovani	811	0,770
	Exit		811	0,940

IPD: Individualni podaci pacijenta; DBS: Duboka moždana stimulacija; ECT:

Elektrokonvulzivna terapija; LOCF: Poslednja opservacija prenesena unapred; J0-J28:

Dan 0-Dan 28; rTMS: Repetitivna transkranijalna magnetna stimulacija.

Korelacije između skala MADRS i BDI

Studija	Vreme ocenjivanja	Poreklo koeficijenata korelacije	Veličina uzorka	Koeficijent korelacije
Hosang i sar., 2012	Baseline	Publikovani	96	0,880
	Follow-up		96	0,890
Lempérière i sar., 1984	J0	Publikovani	75	0,660
	J7		73	0,760
	J14		74	0,870
	J28		74	0,910
Luborzewski i sar., 2007	Pre-rTMS	IPD	17	0,691
	Post-rTMS		17	0,938
Małyszczak i sar., 2006	Before treatment	IPD	6	0,951
	After treatment		6	0,566
Martinsen i sar., 1995	Admission	Publikovani	117	0,670
	Discharge		92	0,800
Merkl i sar., 2013	Baseline	IPD	6	0,115
	Month 6		6	0,884
Palmio i sar., 2010	Pre-ECT	IPD	10	0,600
	Post-ECT		10	0,725
Rotter i sar., 2013	Pre-ECT	IPD	14	0,619
	Post-ECT		14	0,798
Uher i sar., 2012	Baseline	Publikovani	811	0,580
	Exit		811	0,850

IPD: Individualni podaci pacijenta; DBS: Duboka moždana stimulacija; ECT:

Elektrokonvulzivna terapija; LOCF: Poslednja opservacija prenesena unapred; J0-J28:

Dan 0-Dan 28; rTMS: Repetitivna transkranijalna magnetna stimulacija.

Prilog 7 – Studije u kojima su saopštene ocene dijagnostičke valjanosti

Dijagnostička valjanost za HAMD skalu

Studija	Time	n	Sn	Sp	PPV	NPV	AUC	LR+	LR-	DOR
Berg i sar., 2009	Acute phase	90	0,80	0,93	0,40	0,99		11,33	0,21	52,71
	6 month	91	0,50	0,94	0,44	0,95		8,30	0,53	15,6
Kang i sar., 2013	At 2 week	423	0,89	0,84	0,34	0,99	0,92	5,64	0,13	42,72
	At 1 year	288	0,96	0,89	0,44	1,00	0,96	8,72	0,05	185,62
Thompson i sar., 1998	8 week	108	0,99	0,80	0,98	0,89		4,95	0,01	388
	12 week	141	1,00	0,87	0,98	1,00		7,50	0,00	-
	20 week	135	0,99	0,75	0,98	0,90		3,97	0,01	366
	28 week	135	1,00	0,69	0,97	1,00		3,25	0,00	-
Weiss i sar., 1989	Admission	149	0,85	0,49	0,14	0,97		1,67	0,31	5,34
	2 weeks	149	0,69	0,79	0,24	0,96		3,25	0,39	8,30
	4 weeks	149	0,38	0,89	0,25	0,94		3,49	0,69	5,04

Studija	Time	Primarna Dg	Status depresivnosti prema kriterijumskom standardu	Dg prag	TP/ (TP+FN)	TN/ (TN+FP)	Kriterijumski standard
Berg i sar., 2009	Acute phase	Moždani udar	Major depresija / nedeprisivni	10	4/5	79/85	DSM-III-R
	6 month				4/8	78/83	
Kang i sar., 2013	At 2 week	Moždani udar	Major depresija / nedeprisivni	12	32/36	326/387	DSM-IV
	At 1 year						
Thompson i sar., 1998	8 week	Postaparatalni period	Depresija / nedeprisivni		8/10	97/98	RDC
	12 week				13/15	126/126	
	20 week				9/12	122/123	
	28 week				9/13	122/122	
Weiss i sar., 1989	Admission	Zavisnost od kokaina	Major depresija / nedeprisivni	17	11/13	67/136	DSM-III
	2 weeks				9/13	107/136	
	4 weeks				5/13	121/136	

Dijagnostička valjanost za MADRS skalu

Studija	Time	n	Sn	Sp	PPV	NPV	AUC	LR+	LR-	DOR
Kang i sar., 2013	At 2 week	423	0,89	0,79	0,28	0,99		4,26	0,14	30,22
	At 1 year	288	0,90	0,84	0,34	0,99		5,76	0,10	58,14
Laska i sar., 2007	Baseline	73	0,67	0,93	0,71	0,91		9,67	0,36	27,00
	6 months	77	0,67	0,97	0,67	0,97		23,67	0,34	69,00

Studija	Time	Primarna Dg	Status depresivnosti prema kriterijumskom standardu	Dg prag	TP/ (TP+FN)	TN/ (TN+FP)	Kriterijumski standard
Kang i sar., 2013	At 2 week	Moždani udar	Major depresija / nedeprisivni	8	32/36	81/306	DSM IV
	At 1 year				22/24	42/222	
Laska i sar., 2007	Baseline	Moždani udar		10	10/15	54/58	DSM IV
	6 months				4/6	69/71	

Prilog 8 – Studije u kojima su saopštene ocene faktorske valjanosti

Faktorska valjanost za HAMD

Studija	vreme	n	Broj ekstrahovanih faktora	Ukupna objašnjena varijansa (%)	Naziv prvog faktora	Varijansa objasnjena prvim ekstrahovanim faktorom (%)
Addington i sar., 1996	Time 1	112	7	65,5		21,3
	Time 2	89	7	71,8		26,5
Bent-Hansen i sar., 2003	Baseline	230	6			12,4
	Week 3	204	6			22,9
Licht i sar., 2005	Baseline	1629	7			14,6
	2 weeks	1581	6			17,5
	4 weeks	1459	5			27,8
Maier i sar., 1985	P1	107	4	22,1	Inhibicija	
	P3	88	4	33,1		
Schennach i sar., 2012	Admision	278	3	40,0		14,0
	Discharge	278	3	42,0		16,0
Steinmeyer i Möller, 1992	Baseline	223	6			
	Endpoint	174	2			

Struktura faktora po ajtemima na prvoj i poslednjoj proceni za HAMD skalu

HAMD ajtemi	Addington, 1996		Maier, 1985		Schennah, 2012		Steinmeyer, 1992	
	Time 1	Time 2	P1	P2	admission	discharge	Baseline	Endpoint
1 Depresivno raspoloženje	+	+	+			+		+
2 Osećanje krivice	+							
3 Suicid			+					+
4 Rana insomnija								+
5 Teškoće prosnivanja								+
6 Kasna insomnija								+
7 Rad i aktivnost	+			+	+	+		+
8 Retardacija			+	+	+	+		
9 Agitacija							+	
10 Anksioznost		+						+
11 Telesna anksioznost	+	+			+		+	+
12 Telesni gastrointestinalni simptomi		+	+			+		+
13 Opšti telesni simptomi	+			+	+	+		+
14 Genitalni simptomi								
15 Hipohondrijaza					+		+	
16 Uvid u bolest		+						
17 Gubitak telesne težine	+	+		+				
Objašnjena varijansa	21,3	26,5			0,14	0,16		
Broj ajtema u prvom faktoru	6	6	4	4	5	5	3	10

Faktorska valjanost za MADRS skalu

Studija	vreme	n	Broj ekstra faktora	Ukupna objašnjena varijansa (%)	Naziv prvog faktora	Varijansa objasnjena prvim ekstrahovanim faktorom (%)
Bent-Hansen i sar., 2003	Baseline	230	5			17,1
	Week 3	204	4			27,7
Bondolfi i sar., 2010	Time 1 (Study inclusion)	63	2			42,3
	Time 2 (4 weeks later)	63	2			57,6
Galinowski i Lehert, 1995	Day 0	137	3		Opšti faktor	32,3
	Day 28	137	1			66,3
Rocca i sar., 2002	Baseline	118	2		Opšti faktor	32,9
	After 8 weeks	100	1			37,0

Spisak skraćenica

AS	Aritmetička sredina
AUC	<i>Area under curve</i> (engl.), Površina ispod krive maksimalne verodostojnosti
BDI	<i>Beck Depression Inventory</i> (engl.), Bekova skala za procenu depresivnosti
CI	<i>Confidence interval</i> (emgl.), Interval poverenja
CVI	Cerebrovaskularni insult
DOR	<i>Diagnostic odds ratio</i> (engl.), Dijagnostički odnos šansi
DSM	<i>Diagnostic and Statistical Manual of Mental Disorders</i> (engl.), Dijagnostičko i statističko uputstvo mentalnih poremećaja
HAMD	<i>Hamilton Depression Rating Scale</i> (engl.), Hamiltonova skala za procenu depresivnosti
ICC	<i>Intraclass correlation coefficient</i> (engl.), Intraklasni koeficijent korelacije
ICD-10	<i>International Classification of Diseases</i> (engl.), Međunarodna klasifikacija bolesti, deseta revizija
LN	Lažno negativni u dijagnostičkoj tabeli 2×2
LP	Lažno pozitivni u dijagnostičkoj tabeli 2×2
LR+	<i>Likelihood ratio for a positive result</i> (engl.), Odnos verodostojnosti pozitivnog testa
LR-	<i>Likelihood ratio for a negative result</i> (engl.), Odnos verodostojnosti negativnog testa

MADRS	<i>Montgomery-Åsberg Depression Rating Scale</i> (engl.), Mongomeri-Åsbergova skala za procenu depresivnosti
NPV	<i>Negative predictive value</i> (engl.), Negativna prediktivna vrednost
PPV	<i>Positive predictive value</i> (engl.), Pozitivna prediktivna vrednost
RDC	<i>Research Diagnostic Criteria</i> (engl.), Istraživački dijagnostički kriterijumi
RVE	<i>Robust Variance Estimation</i> (engl.), Robusna ocena varijanse
SD	Standardna devijacija
Se	Senzitivnost
Sp	Specifičnost
SROC	<i>Summary receiver operating characteristics</i> (engl.), Sumarna kriva maksimalne verodostojnosti
TN	Tačno negativni u dijagnostičkoj tabeli 2×2
TP	Tačno pozitivni u dijagnostičkoj tabeli 2×2

Biografija

Zoran (Milete) Bukumirić rođen je 21. marta 1973. godine u Peći. Osnovnu školu završio je u Goraždevcu a srednju u Peći (Gimnazija, prirodno-matematički smer) sa odličnim uspehom i Vukovom diplomom. Zvanje doktora medicine stekao je na Medicinskom fakultetu Priština u decembru 2000. godine sa prosečnom ocenom 9,08. U toku studiranja učestvovao je u radu više studentskih Kongresa i bio aktivan u studentskim organizacijama. Naziv magistra nauka dobio je na Medicinskom Fakultetu Priština-Kosovska Mitrovica marta 2004. godine, odbranom rada pod nazivom: „Efekti esomeprazola i diazepama na razvoj stres ulkusa kod eksperimentalnih životinja“. Specijalistički ispit iz Kliničke farmakologije, položio je 2009. godine na Zavodu za farmakologiju, toksikologiju i kliničku farmakologiju, Medicinskog fakuteta, Univerziteta u Novom Sadu. Na Medicinskom fakultetu Univerziteta u Prištini radio je ako asistent na: Institutu za farmakologiju i toksikologiju (2001-2008. godine) i Institutu za medicinsku statistiku i informatiku (2008-2013. godine). Radi kao asistent na Institutu za medicinsku statistiku i informatiku, Medicinskog fakuteta Univerziteta u Beogradu.

Prilog 1.

Potpisani-a Zoran Bukumirić

broj indeksa _____

Izjavljujem

da je doktorska disertacija pod naslovom

„Sistematski pregled sa meta-analizom pouzdanosti i valjanosti psihijatrijskih skala u longitudinalnim studijama”

- rezultat sopstvenog istraživačkog rada,
- da predložena disertacija u celini ni u delovima nije bila predložena za dobijanje bilo koje diplome prema studijskim programima drugih visokoškolskih ustanova,
- da su rezultati korektno navedeni i
- da nisam kršio/la autorska prava i koristio intelektualnu svojinu drugih lica.

Potpis doktoranda

U Beogradu, 05.04.2016. godine



Prilog 2.

Izjava o istovetnosti štampane i elektronske verzije doktorskog rada

Ime i prezime autora Zoran Bukumirić

Broj indeksa _____

Studijski program _____

Naslov rada „Sistematski pregled sa meta-analizom pouzdanosti i valjanosti
psihijatrijskih skala u longitudinalnim studijama”

Mentor Prof. dr Goran Trajković

Potpisani/a Zoran Bukumirić

Izjavljujem da je štampana verzija mog doktorskog rada istovetna elektronskoj verziji koju sam predao/la za objavljivanje na portalu **Digitalnog repozitorijuma Univerziteta u Beogradu**.

Dozvoljavam da se objave moji lični podaci vezani za dobijanje akademskog zvanja doktora nauka, kao što su ime i prezime, godina i mesto rođenja i datum odbrane rada.

Ovi lični podaci mogu se objaviti na mrežnim stranicama digitalne biblioteke, u elektronskom katalogu i u publikacijama Univerziteta u Beogradu.

Potpis doktoranda

U Beogradu, 05.04.2016. godine



Prilog 3.

Izjava o korišćenju

Ovlašćujem Univerzitetsku biblioteku „Svetozar Marković“ da u Digitalni repozitorijum Univerziteta u Beogradu unese moju doktorsku disertaciju pod naslovom:

„Sistematski pregled sa meta-analizom pouzdanosti i valjanosti psihijatrijskih
skala u longitudinalnim studijama”

koja je moje autorsko delo.

Disertaciju sa svim priložima predao/la sam u elektronskom formatu pogodnom za trajno arhiviranje.

Moju doktorsku disertaciju pohranjenu u Digitalni repozitorijum Univerziteta u Beogradu mogu da koriste svi koji poštuju odredbe sadržane u odabranom tipu licence Kreativne zajednice (Creative Commons) za koju sam se odlučio/la.

1. Autorstvo
2. Autorstvo - nekomercijalno
3. Autorstvo – nekomercijalno – bez prerade
4. Autorstvo – nekomercijalno – deliti pod istim uslovima
5. Autorstvo – bez prerade
6. Autorstvo – deliti pod istim uslovima

(Molimo da zaokružite samo jednu od šest ponuđenih licenci, kratak opis licenci dat je na poleđini lista).

U Beogradu, 05.04.2016. godine

Potpis doktoranda