



Metodološki pogovori:

Veljavnost in zanesljivost merskega pripomočka

Anja ŽNIDARŠIČ¹, Špela SELAK²

¹ Univerza v Mariboru, Fakulteta za organizacijske vede

² Nacionalni inštitut za javno zdravje

Povzetek: Uporaba veljavnih in zanesljivih merskih pripomočkov predstavlja ključni element kakovosti raziskovanja. Zato je treba upoštevati različne kriterije za zagotavljanje čim veljavnejšega in zanesljivejšega merskega pripomočka pred, med izvedeno raziskavo in po njej. V prispevku s pomočjo didaktične modalitete dialoga med raziskovalcem in metodologom osvetlimo pomembne vidike zagotavljanja in ocenjevanja zanesljivosti in veljavnosti merskega pripomočka, ki naj jim raziskovalec posveča pozornost, ter vprašanja in izzive, s katerimi se ob tem srečuje.



V prvem prispevku Metodoloških pogovorov smo se posvetili načrtovanju in zasnovi raziskave (1). Osredotočili smo se na opredelitev ciljev in raziskovalnih vprašanj, populacije in izbire načina vzorčenja, izvedbi preliminarnih analiz, kot so na primer poglobljeni intervjuji in fokusne skupine, ter se posvetili tudi sestavljanju anketnega vprašalnika in oblikovanja posameznih anketnih vprašanj, skupaj s pripadajočimi odgovori in ne nazadnje tudi motiviranju anketirancev, da bi v čim večji meri odgovorili na anketni vprašalnik.

V pričujočem prispevku pogovor med raziskovalcem in metodologom nadaljujemo in se posvečamo veljavnosti in zanesljivosti merskega pripomočka, v našem primeru vprašalnika, kot dvema glavnima kriterijema dobrega merskega pripomočka. V raziskovanju na področju družbenih in humanističnih ved ter tudi na nekaterih drugih področjih, kot je na primer področje zdravstva, spremenljivke in pojave pogosto merimo posredno preko abstraktnih konceptov oziroma teoretičnih konstruktov. Pri tem uporaba veljavnih in zanesljivih merskih pripomočkov predstavlja ključni element kakovosti raziskovanja, proces razvijanja in validacije merskega pripomočka

Raziskovalec: »Zaključili smo z zbiranjem podatkov. Poleg razširjenosti smo preverjali tudi odnos mladih do drog v Sloveniji. Kako lahko preverim, ali je bilo merjenje zanesljivo in veljavno ter bodo rezultati zares odražali, kar smo želeli raziskati?«

Metodolog: »Kot že veš, moramo o veljavnosti razmišljati že pri sami zasnovi raziskave oziroma natančneje pri načrtovanju anketnega vprašalnika. Veljavnost merjenja ali merskega pripomočka namreč označuje stopnjo, do katere indikator/spremenljivka ustrezno meri teoretični konstrukt, ki naj bi ga merila (5). Veljavnost se torej nanaša na oceno, kako dobro merski pripomoček meri tisto, čemur je namenjen. Da bi dosegli čim višjo stopnjo veljavnosti, moramo čim natančneje in popolneje operacionalizirati teoretične konstrukte, ki jih želimo raziskovati (6).«

Raziskovalec: »Če prav razumem, se veljavnost torej dotika predvsem odnosa med teoretičnimi in merjenimi spremenljivkami, ter se sprašujemo, ali smo res izmerili tisto, kar smo želeli, torej odnos mladih do drog. Kako pa bi to lahko preverili oziroma objektivno utemeljili?«

Metodolog: »Enoznačnega in enostavnega odgovora, kako preveriti veljavnost merskega inštrumenta, žal ni. Glavni razlog za to je, da

pa je v veliki meri osredotočen na odpravljanje oz. zmanjševanje napak v procesu merjenja (2). Zato je treba upoštevati različne kriterije za zagotavljanje čim veljavnejšega in zanesljivejšega merskega pripomočka tako pred in med izvedbo raziskave kot tudi po njej.

Tudi tokrat obstoječi metodološki izziv obravnavamo s pomočjo didaktične modalitete dialoga med raziskovalcem in metodologom (3, 4), s čimer želimo osvetliti elemente zanesljivosti in veljavnosti vprašalnika, ki naj jim raziskovalec posveča pozornost, ter vprašanja in izzive, s katerimi se ob tem srečuje. Tako podrobno opišemo veljavnost in njene različne vidike ter zanesljivost, kar podpremo s praktičnimi primeri. Predstavimo statistične teste in metode za preverjanje oziroma ocenjevanje teh dveh kriterijev dobrega merskega pripomočka. Na veljavnost in zanesljivost merjenja pogledamo tako z vidika kvantitativnih kot tudi kvalitativnih raziskav. Dodatno se dotaknemo tudi nekaterih drugih kriterijev za ocenjevanje kakovosti izpeljane raziskave ter predstavimo možne napake raziskovalcev v procesu razvijanja in validacije merskega pripomočka.

poznamo več vrst veljavnosti (angl. validity), kot so vsebinska veljavnost (angl. content validity), kriterijska veljavnost (angl. criterion-related validity), konstruktna veljavnost (angl. construct validity), notranja veljavnost (angl. internal validity), zunanja veljavnost (angl. external validity) (6), ki jih preverjamo z različnimi nameni.

Raziskovalec: »Ali se vsebinska veljavnost nanaša na vsebino vprašanj?«

Metodolog: »Vsebinska veljavnost se nanaša na to, ali merski pripomoček oziroma vključene merjene spremenljivke zajamejo celoten obseg vsebine raziskave glede na postavljene cilje. V tvojem primeru bi torej to pomenilo, ali vsa vprašanja oziroma merjene spremenljivke, vključene v tvoj merski pripomoček oziroma vprašalnik, zajamejo celotno problematiko odnosa mladih do drog. Ključni problem tega tipa veljavnosti je, da ne moremo objektivno opredeliti stopnje, do katere operacionalizacija predstavlja teoretični konstrukt, oziroma če ga sploh predstavlja ter ali so v operacionalizacijo zajete vse razsežnosti pojma in ali smo katere razsežnosti izpustili/pozabili. Pri tem velja upoštevati dve osnovni vodili za določanje razsežnosti konstrukta, in sicer popoln pregled literature, s katerim lahko ugotovimo, kako so

teoretični konstrukt obravnavali drugi raziskovalci, ter lastno vedenje in opazovanje (6).«

Raziskovalec: »Če prav razumem, je torej dovolj, da rečem, da je merjenje vsebinsko veljavno, če sem pri pregledu literature pazljiv in zajamem v vprašalnik vse razsežnosti posameznega konstrukta?«

Metodolog: »Delno bi lahko tako rekel, saj če raziskovalec sam preverja oziroma utemeljuje veljavnost vprašalnika, temu rečemo zdravorazumska veljavnost (angl. face validity). Sicer pa vse skupaj ni tako enostavno. Raziskovalci (6) pogosto navajajo dve razsežnosti vsebinske veljavnosti, poleg te še vzorčno veljavnost (angl. sampling validity). Zdravorazumska veljavnost je najšibkejši kriterij preverjanja veljavnosti, saj le raziskovalec sam preveri, ali je konstrukt meril to, kar predvideva, da naj bi meril (6). Vzorčna veljavnost pa se nanaša na to, ali sta število in raznolikost merjenih spremenljivk dovolj velika, da zajamemo vse razsežnosti obravnavanega konstrukta (6).«

Raziskovalec: »Prej so bile omenjene tudi druge vrste veljavnosti. Kaj pomeni kriterijska veljavnost?«

Metodolog: »Kriterijska veljavnost se nanaša na korelacijo med merjeno spremenljivko in teoretično spremenljivko (6). V praksi seveda teoretične spremenljivke ne moremo meriti, zato poiščemo ustrezno kriterijsko spremenljivko, za katero vemo, da je močno pozitivno povezana (korelira) s teoretično spremenljivko. Poznamo dve vrsti kriterijske veljavnosti, in sicer napovedno veljavnost (angl. predictive validity) in sočasno veljavnost (angl. concurrent validity).«

Raziskovalec: »Ali to pomeni, da bi lahko torej na podlagi odnosa mladih do alkohola in kajenja napovedali odnos mladih do drog ter tako preverili kriterijsko veljavnost?«

Metodolog: »Tako je, odnos do alkohola in kajenja bi lahko bila tvoja kriterija za preverjanje kriterijske veljavnosti.«

Raziskovalec: »Kaj pa pri napovedni veljavnosti? Ali v tem primeru napovemo, kakšne odgovore pričakujemo?«

Metodolog: »Ne, ne napovemo, kakšne odgovore anketirancev pričakujemo. Napovedna veljavnost odraža možnost merskega pripomočka (oziroma

posamezne merjene spremenljivke), da napove prihodnje dogodke.«

Raziskovalec: »Torej napovedne veljavnosti ne moremo preveriti v času zasnove raziskave, niti ne takoj za tem, ko so podatki zbrani?«

Metodolog: »Ne. Za preverjanje napovedne veljavnosti bi potrebovali neko drugo, kriterijsko spremenljivko, izmerjeno po zaključku naše raziskave. Merski postopek ima torej napovedno veljavnost, če so predvidevanja na podlagi merjene spremenljivke skladna s stanjem posameznika v prihodnosti, ki je izmerjeno s pomočjo ustrezne kriterijske spremenljivke.«

Raziskovalec: »Napovedna veljavnost je torej le en vidik kriterijske veljavnosti, drugi pa je sočasna veljavnost, mar ne? Ali pri tej veljavnosti tudi primerjamo povezanost merjene spremenljivke s kakšno drugo, znano spremenljivko?«

Metodolog: »Tako je. Pri sočasni veljavnosti ugotavljamo stopnjo povezanosti rezultatov merjene spremenljivke z obstoječo kriterijsko spremenljivko v isti časovni točki. Povedano drugače, pri sočasni veljavnosti gre za primerjavo merskega inštrumenta z že obstoječim in preverjenim inštrumentom, in sicer naj bi dobili podobne rezultate kot v primeru uporabe obstoječega, že validiranega vprašalnika. V tvojem primeru bi tako moral dobiti podobne rezultate kot z drugim primerljivim vprašalnikom.«

Raziskovalec: »Kako pa je s konstruktno veljavnostjo, ki je bila omenjena prej med naštevanjem različnih vrst veljavnosti? Sklepam, da v tem primeru, kot nakazuje že ime, ne preverjamo veljavnosti posamezne spremenljivke, temveč skupek vseh spremenljivk, vključenih v teoretični konstrukt. Kako pa preverjamo ta tip veljavnosti?«

Metodolog: »Res je, konstruktna veljavnost opredeljuje, ali se merjeni indikatorji vedejo konsistentno glede na umeščenost v teoretski okvir. Tako morajo biti spremenljivke, ki merijo posamezni konstrukt, med seboj zelo povezane (morajo visoko korelirati), konstrukti pa se med seboj povezujejo (korelirajo) skladno s teoretičnimi predpostavkami. Če se med seboj povezujejo spremenljivke, ki merijo različna konstrukta, to pomeni, da sta med seboj povezana tudi sama konstrukta (6). V tvojem primeru bi to pomenilo, da sta povezana na primer konstrukta odnos do trdih

in odnos do mehkih drog, saj so medsebojno povezane spremenljivke obeh konstruktov.«

Raziskovalec: »Ob branju znanstvenih člankov sem zasledil, da pri utemeljevanju konstruktne veljavnosti raziskovalci navadno omenjajo diskriminantno veljavnost.«

Metodolog: »Tako je, raziskovalci navadno v svojih analizah navajajo diskriminantno in konvergentno veljavnost. Sicer pa metodologiji konstruktno veljavnost preverjajo z več različnih vidikov oziroma jo delijo v več skupin, kot je na primer razčlenitev veljavnosti v šest komponent. Ta opredeljuje teoretično smiselnost konstruktov, empirično smiselnost konstruktov, notranjo konsistentnost posameznih teoretičnih spremenljivk, konvergentno veljavnost, diskriminantno (tudi diskriminacijsko) veljavnost in nomološko veljavnost (6).«

Raziskovalec: »Kaj imaš v mislih s teoretično smiselnostjo konstrukta?«

Metodolog: »Kot pove že ime, gre za to, ali je konstrukt ustrezno izpeljan iz teorije. Konstrukt lahko opredelimo kot zamisel stalnih, običih in nujnih lastnosti in odnosov dane vrste predmetov–odnosov–procesov (6). Za ustrezno teoretično opredelitev konstruktov je treba biti pozoren na naravo in konsistentnost jezika (6), in sicer moramo biti pozorni na pojmovno jasnost in pojmovno nedvoumnost ter neprotislovnost.«

Raziskovalec: »Verjetno moramo biti poleg teoretične smiselnosti konstrukta pozorni tudi na empirično smiselnost konstruktov?«

Metodolog: »Tako je. Empirična smiselnost konstrukta se nanaša na odnos med teoretičnimi spremenljivkami in operacionalizacijo teh, kar omogoča empirično raziskovanje (6).«

Raziskovalec: »Kako pa je z notranjo konsistentnostjo (skladnostjo) posameznih teoretičnih spremenljivk?«

Metodolog: »Notranja konsistentnost (skladnost) posameznih teoretičnih spremenljivk se nanaša na zanesljivost merjenja posameznega konstrukta. To lahko preverjamo z enodimenzionalnostjo posameznega konstrukta, in sicer na primer s faktorsko analizo ali s strukturnim modelom (6). V tvojem primeru bi torej notranjo konsistentnost (skladnost) lahko preverili tako, da bi izvedli

faktorsko analizo, ki bi pokazala en faktor, kar bi bilo razvidno iz tega, da bi bila prva lastna vrednost višja od 1 in hkrati tudi precej višja od druge lastne vrednosti, ki bi bila nižja od 1.«

Raziskovalec: »Na kaj pa se nanašata konvergentna in diskriminantna veljavnost? V raziskovalnih besedilih namreč zasledim, da raziskovalci poleg konvergentne veljavnosti pogosto omenjajo tudi diskriminantno. Kaj to pomeni?«

Metodolog: »Konvergentna veljavnost se nanaša na stopnjo skladnosti dveh ali več poskusov merjenja istega konstrukta z različnimi merskimi postopki (6), s čimer preverimo, ali so različni merski postopki nadomestljivi. Z diskriminantno veljavnostjo pa preverjamo, ali so konstrukti, ki so glede na teorijo nepovezani, tudi dejansko nepovezani. To pomeni, da morajo biti korelacije med spremenljivkami, ki merijo nek drug teoretični konstrukt, nizke.«

Raziskovalec: »Raziskovalci včasih omenjajo tudi pojma notranja in zunanja veljavnost. Kako pa je s tem?«

Metodolog: »Notranja veljavnost se nanaša na zaupanje v ugotovitve raziskave, ki se nanašajo na obravnavani vzorec enot in sam kontekst raziskave (6). Torej govorimo o tem, ali lahko rezultate raziskave pripišemo vzroku, ki smo jih proučevali, in ne nekim zunanjim vzrokom. Na drugi strani pa z zunanjo veljavnostjo ugotavljamo, ali rezultate raziskave lahko posplošimo na populacijo, kontekste in časovna obdobja izven okvirja dane raziskave (6).«

Raziskovalec: »Kako pa bi to lahko preveril v mojem primeru?«

Metodolog: »Zunanjo veljavnost bi lahko preveril na primer tako, da bi svojo raziskavo o odnosu do drog izvedel v nekem drugem kulturnem okolju oziroma v drugi državi in bi seveda želel dobiti podobne rezultate.«

Raziskovalec: »Zanima me še, kaj pomeni nomološka veljavnost?«

Metodolog: »Nomološka veljavnost pa izraža stopnjo, do katere so proučevani konstrukti, ki naj bi bili povezani v teoriji, povezani tudi empirično (6).«

Raziskovalec: »Vidim, da obstaja resnično veliko definicij veljavnosti, kar raziskovalca lahko hitro zmede.«

Metodolog: »Res je, metodologi navajajo različne klasifikacije oziroma razčlenbe komponent veljavnosti, kar lahko deluje malce konfuzno. Zasedimo tudi ekološko veljavnost, ki se nanaša na to, v kolikšni meri lahko rezultate raziskave apliciramo v vsakdanje življenje ljudi (7).«

Raziskovalec: »Kako pa vse to poteka v praksi? Kako lahko preverimo veljavnost vprašalnika? Katere statistične teste in metode moramo uporabiti?«

Metodolog: »Za ugotavljanje veljavnosti konstrukta (6) lahko uporabljamo različne multivariatne pristope (analiza glavnih komponent, (eksploratorna in konfirmatorna) faktorska analiza, kanonična korelacijska analiza) ali pristop MTMM (večrazsežnostni – večmetodni pristop, angl. multitrait-multimethod approach).«

Raziskovalec: »V objavljenih raziskavah raziskovalci res velikokrat objavljajo rezultate faktorske analize. Ali gre tu za eksploratorno ali konfirmatorno faktorsko analizo?«

Metodolog: »Eksploratorno faktorsko analizo (angl. exploratory factor analysis) navadno uporabljamo v preliminarnih raziskavah, ko raziskujemo strukturo merjenih spremenljivk, saj konstrukti, v katere spadajo posamezne spremenljivke, še niso definirani oziroma predhodno potrjeni na podlagi podatkov kakšne raziskave. Konfirmatorno faktorsko analizo (angl. confirmatory factor analysis) pa uporabljamo za preverjanje oz. potrjevanje pripadnosti določenih merjenih spremenljivk posameznemu teoretičnemu konstruktju.«

Raziskovalec: »Torej, če bi želeli preveriti veljavnost konstrukta odnosa mladih do drog, bi morali uporabiti konfirmatorno faktorsko analizo?«

Metodolog: »Tako je, veljavnost konstrukta lahko preverjamo tudi s konfirmatorno faktorsko analizo. Za ugotavljanje konvergentne veljavnosti morajo spremenljivke posameznega konstrukta visoko korelirati med seboj. Konvergentno veljavnost raziščemo na podlagi treh kazalnikov (8). Ocene standardiziranih faktorskih uteži naj bi bile višje od 0,5 (ali celo od 0,7), kompozitna zanesljivost (angl. composite reliability – CR) za vsak konstrukt naj bi

preseгла 0,7 ter povprečna izražena varianca (angl. average variance extracted – AVE) za vsak konstrukt naj bi bila višja kot 0,5. Diskriminantno veljavnost pa preverjamo na drugačen način. Kadar na primer preverjamo odvisnosti med konstrukti s pomočjo strukturnega modeliranja (angl. structural equation modeling) v prvem koraku (9), na podlagi izvedene konfirmatorne faktorske analize diskriminantno veljavnost navadno preverjamo s t. i. Fornell-Larckerjevim kriterijem (10), ki pravi, da morajo biti vrednosti kvadratnega korena iz AVE (angl. average variance extracted) višje kot pripadajoče korelacije med ostalimi konstrukti oziroma latentnimi spremenljivkami.«

Raziskovalec: »Poleg veljavnosti se pogosto navaja tudi pojem zanesljivosti vprašalnika. Kaj točno torej pomeni zanesljivost?«

Metodolog: »Zanesljivost (angl. reliability) merskega pripomočka oziroma merjenja posameznega teoretičnega konstrukta je opredeljena s ponovljivostjo merjenja, kar pomeni, da moramo v primeru, da merjenje ponovimo v enakih ali zelo podobnih razmerah, dobiti enake ali vsaj zelo podobne rezultate (5,6).«

Raziskovalec: »Na podlagi česa pa lahko ocenimo zanesljivost merjenja?«

Metodolog: »Zanesljivost merjenja ocenjujemo z različnimi metodami, ki jih delimo v dve skupini (6). To so metode, ki temeljijo na ponovitvi postopka merjenja na istih enotah v dveh časovnih točkah (npr. metoda retesta, metoda alternativne oblike) ter metode, ki temeljijo na merjenju dejanske spremenljivke z več enakovrednimi spremenljivkami v isti časovni točki (npr. metoda razpolovitve, metoda notranje konsistentnosti oziroma skladnosti). V literaturi najpogosteje zasledimo metodo notranje konsistentnosti (skladnosti), ki temelji na izračunu kovarianc oziroma korelacij med spremenljivkami, ki merijo posamezen konstrukt. Rezultat je izražen s koeficientom, ki ga poznamo pod imenom Cronbachov (alfa) koeficient zanesljivosti.«

Raziskovalec: »Torej, če prav razumem, bi bilo tudi v našem primeru najbolje izračunati Cronbachov koeficient alfa za celoten vprašalnik? Kakšne vrednosti bodo nakazovale ustrezno visoko zanesljivost?«

Metodolog: »Ta sklep ni povsem pravilen. Cronbachov koeficient alfa navadno izračunamo za

vsak teoretični konstrukt posebej in ne le za celoten vprašalnik. Navadno se kot meja, ki nakazuje, da je merjenje posameznega konstrukta zanesljivo, navaja 0.8 ali 0.7 (11). Treba pa je opozoriti, da je vrednost koeficienta odvisna tudi od števila merjenih spremenljivk, zato lahko ob velikem številu spremenljivk dobimo visoke vrednosti koeficienta, kar ne kaže nujno visoke zanesljivosti merjenja in obratno, pri majhnem številu merjenih spremenljivk lahko dobimo nižje vrednosti Cronbachovega koeficienta alfa.«

Raziskovalec: »Kako pa potem lahko kar najzanesljiveje ugotovim, kakšna je zanesljivost našega merskega pripomočka?«

Metodolog: »V tem primeru ti težko odgovorim enoznačno. Kljub temu da izračunaš Cronbachov koeficient alfa, ni priporočljivo strogo upoštevati le njegove vrednosti, ampak je treba v primeru nizkih vrednosti pogledati, ali je morda vzrok za slednje majhno število spremenljivk.«

Raziskovalec: »Ali je poleg veljavnosti in zanesljivosti še kaj, kar označuje, kako dobro je načrtovana oziroma izpeljana raziskava?«

Metodolog: »Da, med drugim naj bi pri vrednotenju raziskave poleg veljavnosti in zanesljivosti preverili oziroma navedli še ponovljivost (angl. replicability) in posplošljivost (angl. generalizability) (7). Pri tem se ponovljivost nanaša na to, ali lahko raziskavo ponovimo in dobimo tudi podobne rezultate, medtem ko se posplošljivost nanaša na to, ali lahko rezultate posplošimo na druge lokacije oziroma situacije. Ta definicija posplošljivosti je torej skladna s konceptom prej omenjene zunanje veljavnosti.«

Raziskovalec: »Meniš, da je to pomembno tudi v našem primeru?«

Metodolog: »Seveda, pri vsaki raziskavi ne glede na temo je pomembno, da je dobro dokumentirana, tako da jo lahko ponovimo čez leta ali v drugih okoljih ter tako preverimo, ali lahko ugotovitve posplošimo izven okvirov naših raziskave.«

Raziskovalec: »Sedaj sva opisovala, kako v kvantitativnih raziskavah preverjamo, ali je merjenje veljavno in zanesljivo, kar poenostavljeno pomeni, da merimo, to kar smo želeli meriti, in da ob ponovitvi merjenja v podobnih razmerah dobimo podobne rezultate. Kako pa veljavnost in

zanesljivost opredelimo in preverjamo v kvalitativnih študijah?«

Metodolog: »V kvalitativnih študijah namesto o zanesljivosti in veljavnosti govorimo o podatkih, vrednih zaupanja oziroma zanesljivih podatkih (angl. trustworthy data). Raziskovalci (12) se osredotočajo na kredibilnost (angl. credibility), zanesljivost (angl. dependability), potrdljivost (angl. confirmability) in prenosljivost (angl. transferability).«

Raziskovalec: »Kaj pa imaš v mislih s kredibilnostjo?«

Metodolog: »Kredibilnost (v kvantitativnih raziskavah je ustrezen pojem veljavnosti) se nanaša na zaupanje v resničnost ugotovitev, torej ali se rezultati logično povezujejo med seboj. Zanesljivost kvalitativnih podatkov se ne nanaša toliko na samo ponovljivost rezultatov (kot v kvalitativnih raziskavah), temveč na ponovljivost in konsistentnost samega procesa zbiranja podatkov. Povedano drugače, pri kvalitativno zbranih podatkih ne pričakujemo enakih odgovorov, zanesljivost oziroma kredibilnost podatkov se nanaša na to, da bomo dobili logične dosledne vzorce odgovorov, ki so razmeroma stabilni skozi čas.«

Raziskovalec: »Razumem. Kaj pa pomeni potrdljivost?«

Metodolog: »Ta se nanaša na ohranjanje distance med raziskovalcem in opazovanim, torej pomeni, da podatki kar se da natančno odražajo udeleženčevo mnenje, stališča in izkušnje, kar zagotavlja tudi objektivnost merjenja. Skladno s tem mora raziskovalec tudi objektivno dokumentirati svojo vlogo v raziskovalnem procesu, tako predpostavke, pristranskosti in reakcije, ki bi lahko vplivale tako na samo zbiranje podatkov, kot tudi njihovo interpretacijo.«

Raziskovalec: »Sklepam, da se termin prenosljivost verjetno nanaša na prenos ugotovitev na podlagi kvalitativno zbranih podatkov na podobne situacije?«

Metodolog: »Res je, prenosljivost se nanaša na posploševanje. V kvantitativnih raziskavah je cilj posploševanje rezultatov z vzorca na celotno populacijo, medtem ko prenosljivost v kvalitativnih raziskavah pomeni prenosljivost ugotovitev v druge, podobne kontekste, seveda ob predpostavki, da so bili udeleženci raziskave izbrani na način, da je bilo

možno zbrati (vse) različne poglede na raziskovalni problem.«

Raziskovalec: »Ugotavljam, da v raziskovalnem procesu lahko hitro naredimo številne napake. Katere so najpogostejše napake raziskovalcev, na katere naj bomo pozorni?«

Metodolog: »Tako je. Zato se tudi proces razvijanja in validacije merskega pripomočka v veliki meri osredotoča na zmanjševanje napake v procesu merjenja (2). Preden odgovorim na tvoje vprašanje, naj najprej pojasnim, kaj v raziskovanju predstavlja napaka. Po eni od opredelitev (13) napaka predstavlja vsako odstopanje parametrov oziroma rezultatov, zbranih na podlagi vzorca glede na dejanske, prave parametre populacije, kar naj bi se odražalo v štirih tipih napak v (anketnih) raziskavah. Govori o napaki pokritja (angl. coverage error), napaki neodgovora (angl. non-response), vzorčni napaki (angl. sampling error) ter merskih napakah (angl. measurement error).«

Raziskovalec: »Napaka pokritja se verjetno nanaša na to, da ne moremo anketirati vseh ustreznih enot?«

Metodolog: »Do napake pokritja pride, ker nekatere enote niso del vzorčnega načrta. Spomnimo se, da je vzorčni načrt v bistvu spisec enot populacije. In če je vzorčni načrt nepopoln, je delu enot iz populacije posledično odvzeta možnost, da bi bile izbrane v vzorec.«

Raziskovalec: »Omenil si tudi napako neodgovora. Ali v tem primeru anketiranec odkloni sodelovanje v anketi?«

Metodolog: »Pravilno sklepaš. To, da anketiranec zavrne sodelovanje v anketi, je en tip neodgovora. Napaka neodgovora v splošnem nastane zato, ker niso zbrani vsi podatki od vseh enot v vzorcu. Torej anketiranec lahko ne odgovori na celo anketo ali le na del vprašanj.«

Raziskovalec: »Kaj pa vzorčna napaka? Ali je ta posledica slabo izbranega vzorca?«

Metodolog: »Vzorčna napaka nastane takrat, ko je raziskava izvedena na vzorcu in ne na celotni populaciji. Tako vzorčna napaka ni posledica slabo izbranega vzorca, temveč je prisotna vsakič, ko sklepamo o parametru populacije na podlagi vzorca.«

Raziskovalec: »Aha, razumem. Kako pa je z mersko napako? Ali ni to vsaka napaka v procesu merjenja?«

Metodolog: »Merska napaka je razlika med izmerjeno in pravo vrednostjo. Merske napake nastanejo zaradi netočnosti v zabeleženih odgovorih glede na pravo vrednost. Delimo jih na vpliv intervjuarja na odgovore anketiranca, napake anketirancev (nezmožnost podati odgovor, pomanjkanje truda ali informacij, da bi podali točen odgovor ...), napake zaradi oblikovanja oziroma formuliranja vprašanja in vprašalnika ter napake zaradi same metode zbiranja podatkov. Metodologi definirajo tudi skupno napako pri anketiranju (angl. the total survey error), ki zajema vse potencialne vire napak, ki se pojavijo med planiranjem raziskave oziroma ankete ter poročanjem o dobljenih rezultatih (14). Na tem mestu lahko dodamo, da do napak prihaja tudi zaradi slepega zaupanja v obstoječe metode in posledično zanemarjivega preverjanja njihove smiselnosti, uporabnosti in točnosti, tudi zanemarjanja izsledkov primarnih virov literature in uporaba podatkov, predstavljenih v sekundarnih virih, ter na splošno pomanjkanje razumevanja metodoloških orodij in pristopov (15).«

Raziskovalec: »Torej se napake lahko pojavijo na različnih korakih raziskave, lahko jih torej naredi raziskovalec ali pa anketiranec?«

Metodolog: »Prav imaš, napake se lahko pojavijo v kateri koli fazi raziskovanja, lahko bi jih na grobo razdelili tudi na napake v fazi načrtovanja raziskave ter na napake pri analiziranju podatkov. Prva skupina napak tako zajema na primer napake pri načrtovanju raziskave, operacionalizaciji spremenljivk, izbiri vzorca in izvedbi raziskave in te napake je navadno zelo težko odkriti. Druga skupina napak, torej napake pri analiziranju podatkov, pa se lahko pojavijo že pri sami pripravi ali kodiranju podatkov in se kažejo kot napačna izbira prikazanih statistik ali neustrezni grafični prikazi (glede na nivo merjenja spremenljivke), napačna izbira statističnih testov oziroma metod, identifikacija in obravnava osamelcev v podatkih, obravnava manjkajočih vrednosti, napačna interpretacija rezultatov testov (npr. korelacija ne pomeni vpliva) ipd. Prav je, da poudarimo, da se napake v anketah ne pojavijo le zaradi raziskovalcev in pomanjkljivega/napačnega načrtovanja raziskave, temveč pomemben vir napak predstavljajo tudi anketiranci sami.«

Raziskovalec: »Odlično. Hvala za vse odgovore. Nedvomno bodo pripomogli k zagotavljanju in

ocenjevanju veljavnosti našega merskega pripomočka.«

LITERATURA

1. Žnidaršič A, Selak Š. Metodološki pogovori: Izvedba raziskave od postavitve raziskovalnega problema in ciljev, vzorčenja do sestave vprašalnika in izvedbe anketiranja. *Javno Zdr* 2018;2(1):60–71.
2. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm* 2008;65(23):2276–84.
3. Stang A, Poole C. The researcher and the consultant: A dialogue on null hypothesis significance testing. *Eur J Epidemiol* 2013;28(12):939–44.
4. Hamra GB, Stang A, Poole C. The researcher and the consultant: from testing to probability statements. *Eur J Epidemiol* 2015;30(9):1003–8.
5. Jupp V. *The SAGE dictionary of social research methods*. London: SAGE Publications, 2006.
6. Ferligoj A, Leskošek K, Kogovšek T. *Zanesljivost in veljavnost merjenja*. Ljubljana: Fakulteta za družbene vede, 1995.
7. Walliman N. *Social research methods*. London: SAGE publications, 2006.
8. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Errors. *J Mark Res* 1981;18(1):39–5.
9. Schumacker RE, Lomax RG. *A Beginner's Guide to Structural Equation Modeling*. 3rd ed. New York: Routledge, 2010.
10. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *J Mark Res* 1981;18(1):39–50.
11. Field A. *Discovering statistics using IBM SPSS statistics*. 4th ed. London: Sage Publications, 2013.
12. Ulin PR, Robinson ET, Tolley EE. *Qualitative Methods in Public Health. A Field Guide for Applied Research*. San Francisco: Jossey-Bass Publishers, 2005.
13. Groves RM. *Survey errors and survey costs*. New Jersey: John Willey & Sons, 2004.
14. Biemer PP, Lyberg LE. *Introduction to survey quality*. New Jersey: John Willey & Sons, 2003.
15. Ioannidis JPA. Massive citations to misleading methods and research tools: Matthew effect, quotation error and citation copying. *Eur J Epidemiol* 2018;33(11):1021–3.