



Žiga Kozinc¹,
Jure Žitnik¹

Za boljšo prihodnost raziskav na področju športa in kineziologije – poziv k odpravi metodoloških pomanjkljivosti in h krepitvi znanstvene integritete

Izvleček

Znanost o športu in z njo kineziologija se srečujeta s pomembnimi metodološkimi izzivi, ki vplivajo tudi na kakovost in integriteto raziskav. V članku obravnavamo nekatere pomembne metodološke pomanjkljivosti in neetične prakse, kot so subjektivnost pri ocenjevanju kakovosti znanstvenih del, pristranskost pri objavah, drobljenje objav na najmanjšo še objavljivo enoto (angl. salami slicing), post hoc oziroma naknadno postavljanje hipotez, torej po tem, ko so rezultati že znani (angl. HARKing) ter selektivno poročanje in manipuliranje s podatki z namenom potrditve statistično značilne vrednosti (angl. p-hacking). Povzemamo tudi pogoste nepravilnosti v statističnih metodah, kot sta neupoštevanje kovariat in neustrezen izračun velikosti vzorca. Ob tem poudarjamo pomen transparentnosti in dostopnosti podatkov ter vnaprejšnje registracije raziskav za izboljšanje znanstvenih praks. Priporočamo večje sodelovanje s strokovnjaki za statistiko ter uporabo relevantnih smernic za načrtovanje raziskav in poročanje o njihovih izsledkih. Z upoštevanjem vsega naštetega bodo tudi raziskovalci na področju športa in kineziologije v Sloveniji pomembno pripomogli k zvišanju kakovosti raziskav in okrepitvi znanstvene integritete.

Glavne besede: znanstvena integriteta, pristranskost, selektivno poročanje, transparentnost, kakovost raziskav



For a Better Future of Sport Science and Kinesiology Research: A Call to Address Methodological Shortcomings and Strengthen Scientific Integrity

Abstract

Sport science and kinesiology face significant methodological challenges that impact the quality and integrity of research. In the article, we discuss some significant methodological shortcomings and unethical practices, such as subjectivity in assessing the quality of scientific works, publication bias, splitting publications into the smallest publishable unit (known as salami slicing), post hoc hypothesis formulation, that is, after the results are already known (referred to as HARKing), and selective reporting and data manipulation to confirm statistically significant values (known as p-hacking). We also summarize common deficiencies in statistical methods, including the neglect of covariates and inadequate sample size calculation. We emphasize the importance of transparency and data accessibility, as well as the pre-registration of research to improve scientific practices. We recommend increased collaboration with statisticians and the use of relevant guidelines for research planning and reporting. By focusing on these aspects, researchers in the fields of sport and kinesiology in Slovenia can contribute to the improvement of research quality and scientific integrity.

Keywords: scientific integrity, bias, selective reporting, transparency, research quality.

¹Univerza na Primorskem, Fakulteta za vede o zdravju, Izola

■ Uvod

Besedna zveza »objavi ali propadi« (angl. publish or perish) se nanaša na nepisano pravilo v akademskem svetu, ki poudarja pomembnost objavljanih raziskovalnih del za napredovanje in uveljavljanje raziskovalcev (Rawat in Meena, 2014). V akademskem okolju je zaznati nenehen pritisk na visokošolske učitelje in raziskovalce k rednemu objavljaniu ugotovitev v uglednih znanstvenih revijah. Tudi v Sloveniji je napredovanje v višje pedagoške in raziskovalne nazive odvisno od bibliografije, praviloma predvsem od števila in kakovosti objav znanstvenih člankov. Kakovost oziroma pomembnost znanstvenih del je težko merljiv koncept (Aksnes idr., 2019), v večini primerov se lahko oceni šele po določenem obdobju od njihove objave (npr. upoštevajoč število citatov). Pri trenutnem sistemu ocenjevanja znanstvenoraziskovalnega dela ima količina prednost pred kakovostjo, pri čemer se zdi, da potreba po pogostem objavljaniu vodi v nižjo kakovost del. Tako kot v znanosti na splošno (Landhuis, 2016) tudi na področju športa in kineziologije opažamo veliko rast števila znanstvenih objav – med letoma 2000 in 2021 se je število aktivnih revij na tem področju (podatek iz baze Scopus) povečalo s 83 na 121, število objavljenih člankov pa s približno 7.500 na 18.500 letno (Tiller in Ekkekakis, 2023). Razlogov za okrepitev znanstvene produkcije je najverjetneje več, zagotovo med drugim narašča razpoložljivost uporabnih podatkov (Abt idr., 2022; Robertson, 2020). Po drugi strani gre omenjeni porast vsaj deloma pripisati vse večjim pritiskom k objavljaniu znanstvenih člankov.

Raziskave kažejo, da pritiski k objavljaniu znanstvenih del vodijo v nižjo kakovost raziskav in slabe prakse raziskovalcev. Te se lahko pojavijo pri vseh fazah raziskovalnega dela, vse od zasnove raziskave do poročanja o rezultatih in odločanja o objavi (Tiller in Ekkekakis, 2023). Pritisk k objavljaniu je bil prepoznan kot statistično značilen napovedovalec slabih oziroma napačnih praks (Gopalakrishna idr., 2022; Maggio idr., 2019). Po nedavnih ugotovitvah se število umikov že objavljenih člankov v revijah s področja športnih znanosti povečuje (Kardeş idr., 2020). V literaturi je kot odgovor na to mogoče zaslediti vse več pozivov k izboljšanju raziskovalne prakse na področju športne znanosti, športne medicine in kineziologije, od krajših pisem urednikov revij do obsežnejših preglednih člankov (Borg idr., 2023; Büttner idr., 2020; Caldwell idr., 2020;

Sainani idr., 2021; Tiller in Ekkekakis, 2023). Cilj tega prispevka je na podlagi omenjenih pozivov in predlogov za izboljšave raziskovalce na področju športa in kineziologije v Sloveniji seznaniti z najpogostejšimi napakami in slabimi oziroma napačnimi praksami pri znanstvenoraziskovalnem delu ter ponuditi rešitve, ki bodo pripomogle k zvišanju kakovosti raziskav in znanstvene integritete. Pri tem naj poudarimo, da odgovornost ni samo na ramenih raziskovalcev, temveč je to skrb vseh deležnikov v procesu znanstvenoraziskovalnega objavljani (avtorji prispevkov, recenzenti, uredniki revij, založniki in financirji raziskav). V članku obravnavamo izbrane pogoste napake in prakse, predstavljamo pa tudi tiste, ki jih največkrat omenjajo tuji avtorji; članek tako ne predstavlja popisa vseh mogočih napak in slabih praks.

■ Ponarejanje podatkov

Ponarejanje podatkov je najhujša, a k sreči najverjetneje zelo redka praksa v športni znanosti, zato jo bomo le na kratko opisali. V študiji o raziskovalnih praksah več kot 2000 znanstvenikov na večjih ameriških univerzah so poročali, da je 9 % znanstvenikov v preteklosti vsaj nekoliko prirejal podatke (John idr., 2012). Poleg nekaj zelo odmevnih primerov očitnega ponarejanja (Callaway, 2011; Dahlberg in Mahler, 2006) je objektivnih informacij o pogostosti te prakse na področju športa in kineziologije zelo malo. V nedavno objavljenem preglednem članku (Gaspar in Esteves, 2021) z zbiranjem raziskav o slabih praksah je bilo vendarle navedeno, da je ponarejanje podatkov zaznati tudi v športni znanosti, vendar v precej manjšem obsegu kot nekatere druge prakse, kot so plagiatstvo in druge namerne ali namerne slabe prakse, ki jih bomo opisovali v nadaljevanju. Ob tem opozorimo, da pomanjkanje dokazov ne pomeni, da se podatki ne prirejajo. V praksi namreč ni zanesljivih mehanizmov za ugotavljanje, ali so v statistične obdelave vključeni prirejeni podatki. Zato sta integriteta raziskovalca in etično ravnanje temeljnega pomena.

■ Najpogostejše napake in slabe prakse v širšem raziskovalnem procesu

Napake in slabe prakse se lahko pojavijo že pri načrtovanju študije, pozneje pa tudi

pri zbiranju podatkov, statistični analizi in interpretaciji ali poročanju. V tem poglavju se bomo osredotočili na različne oblike pristranskosti pri interpretaciji in poročanju o rezultatih. Nekaj primerov je ponazorjenih na Sliki 1.

Pristranskost pri objavah (»publication bias«)

Pomembna težava v znanstvenih raziskavah je pristranskost pri objavah (angl. publication bias). Ta lahko popači resnično sliko o opravljeni raziskavi glede dane hipoteze ter vpliva na veljavnost pregledov literature in metaanaliz. Pristranskost se pojavi zato, ker je objava raziskovalnih rezultatov odvisna od zanimivosti raziskovalnega vprašanja in pridobljenih ugotovitev (Slika 1). Raziskave s pozitivnimi in statistično značilnimi izidi so objavljene pogostejše kot tiste z negativnimi ali nejasnimi rezultati. Že pred več desetletji so poročali, da na odločitev o (ne)oddaji člankov v recenzijo znanstvenim revijam pomembno vplivajo prav rezultati raziskave (Coursol in Wagner, 1986). Opisani fenomen so poimenovali »problem predala« (angl. file drawer problem). Izraz ponazarja rezultate raziskav, ki končajo zgolj v arhivu in se jih ne pošlje v objavo. Novejša literatura kaže, da je problem opaziti tudi v športni znanosti (Bernards idr., 2017). Analize člankov s področja psihologije športa in vadbe kažejo, da približno 98 % raziskav poroča o vsaj enem statistično značilnem učinku (Twomey idr., 2021), medtem ko je le dobra polovica predhodno registriranih raziskav na področju športne medicine tudi objavljenih (Chahal idr., 2012). Podobno analiza člankov revij s področja športa z najvišjimi faktorji vpliva za leto 2019 poroča, da je 82 % objavljenih člankov potrdilo primarno hipotezo, medtem ko jih je le 18 % poročalo o negativnih (statistično neznačilnih) rezultatih (Büttner idr., 2020). Kot smo že omenili, lahko ta pristranskost vodi do izkrivljenega razumevanja določene tematike oziroma raziskovalnega vprašanja, saj metaanalize vključujejo predvsem raziskave s pozitivnimi izidi, to pa precenjuje resnični učinek intervencije ali zdravljenja. Prepoznavanje in odpravljanje pristranskosti pri objavah je ključno za zagotavljanje natančnosti in integritete znanstvenega znanja. Omeniti je treba, da tovrstna pristranskost ne izvira le iz odločitev raziskovalcev, temveč je razširjena tudi med uredniki revij, saj ti dajejo prednost člankom z novimi, zanimivimi in statistično značilnimi rezultati (Ekmekci, 2017).

Drobljenje objav na najmanjšo še objavljivo enoto »salami slicing«

Izraz rezanje člankov (angl. »salami slicing«) opisuje prakso v znanstvenem svetu, pri kateri raziskovalci ugotovitve ene raziskave razdelijo na več ločenih objav (Xie in Ali, 2023). Ta pristop je pogosto tarča kritik, saj lahko umetno napihne število avtorjevih publikacij, hkrati pa je lahko povezan še z drugimi slabimi praksami, opisanimi v nadaljevanju. Prva težava pri drobljenju člankov je, da lahko razdrobljenost informacij zmanjša celovitost in povezanost znanstvenih spoznanj ter tako oteži razumevanje in sintezo rezultatov pri drugih raziskovalcih (Slika 1). V akademskem svetu je zato taka praksa obravnavana kot vprašanje znanstvene etike, saj lahko vodi v napačno predstavo o obsegu in pomenu posameznikovega raziskovalnega dela (Sasaki in Tan, 2018). Napačno predstavo o ugotovitvah raziskave dobijo tudi bralci. Raziskovalec lahko v študijo učinkovitosti posamezne intervencije vključi veliko spremenljivk, nato pa v enem članku povzame zgolj statistično značilne rezultate in v drugem preostale. Nazadnje je treba izpostaviti še, da lahko t. i. »salami slicing« privede do dvojnega upoštevanja istih podatkov v metaanalizah, s tem pa se popači izračunani skupni učinek. »Salami slicing« se tako pogosto kombinira z drugimi slabimi praksami (opisane so v naslednjih poglavjih), kot so odsotnost statističnih korekcij, »p-hacking« in postavljanje hipotez post hoc. Drobljenje podatkov na več objav je upravičeno le pri obsežnejših raziskavah in projektih (Smart, 2017; Xie in Ali, 2023), vendar je treba tudi v tem primeru za namen transparentnosti to v članku jasno navesti pri opisu metod.

Eksplorativne raziskave in HARKing

V idealnih razmerah bi raziskovalci za vsako hipotezo oziroma raziskovalno vprašanje opravili ločeno potrjevalno raziskavo (angl. confirmatory research). Ta tip raziskave se izvede za potrditev neke hipoteze, znane vnaprej (pred začetkom meritev). Zaradi narave področja pa je velik del študij v športu in kineziologiji eksplorativnih (Twomey idr., 2021). V teh raziskavah se hipoteze postavijo šele po tem, ko so bili podatki že pridobljeni za drug namen. Pri tem gre lahko za sekundarno analizo podatkov predhodne potrjevalne raziskave, skupno analizo podatkov več predhodnih raziskav

ali analizo drugače pridobljenih podatkov (npr. podatkovne baze športnih trenerjev). Prepogosto se dogaja, da raziskovalci izvedejo eksplorativno študijo z že zbranimi podatki, a tega ne navedejo v članku. Ko se eksplorativne raziskave napačno predstavljajo za potrjevalne, se poveča tveganje za netočne, napačne ali neponovljive izide (Ioannidis, 2005); predvsem se občutno poveča možnost za lažno pozitivne rezultate (Begley in Ioannidis, 2015), saj bodo raziskovalci eksplorativne raziskave (še posebej takrat, ko jih lahko enostavno opravijo več) praviloma objavljali le v primeru statistično značilnih rezultatov, potrjevalne raziskave pa ne glede na izid. Poudarjamo, da z izvajanjem eksplorativnih analiz ni nič narobe, nasprotno, so celo zaželene, saj pomagajo usmerjati znanost in lahko vodijo do novih odkritij. Težava nastane pri napačnem predstavljanju eksplorativne raziskave za potrjevalno.

Pričakovani izidi oziroma rezultati intervencijskih študij so običajno raziskovalcem znani vnaprej, na podlagi tega se že pred izvedbo meritev oblikuje hipoteza. Včasih pa se hipoteza ustvari retroaktivno (šele po pridobitvi in analizi podatkov), vendar avtorji v članku predstavijo, kot da je bila hipoteza zasnovana vnaprej. Takšno post hoc oblikovanje hipotez je v angleški literaturi dobilo kratico HARKing (Hypothesizing after the results are known) (Kerr, 1998). Podobno kot predstavljanje eksplorativne študije za potrjevalno HARKing izkrivlja realno sliko, saj ustvarja občutek, da so bili rezultati študije predvidljivi in da so tako bolj zanesljivi (Nosek idr., 2018). Podatkov za področje športa in kineziologije sicer ni, iz drugih ved pa poročajo o zelo visoki prevlenci HARKinga, tudi do 30 % (John idr., 2012).

Selektivno poročanje in »p-hacking«

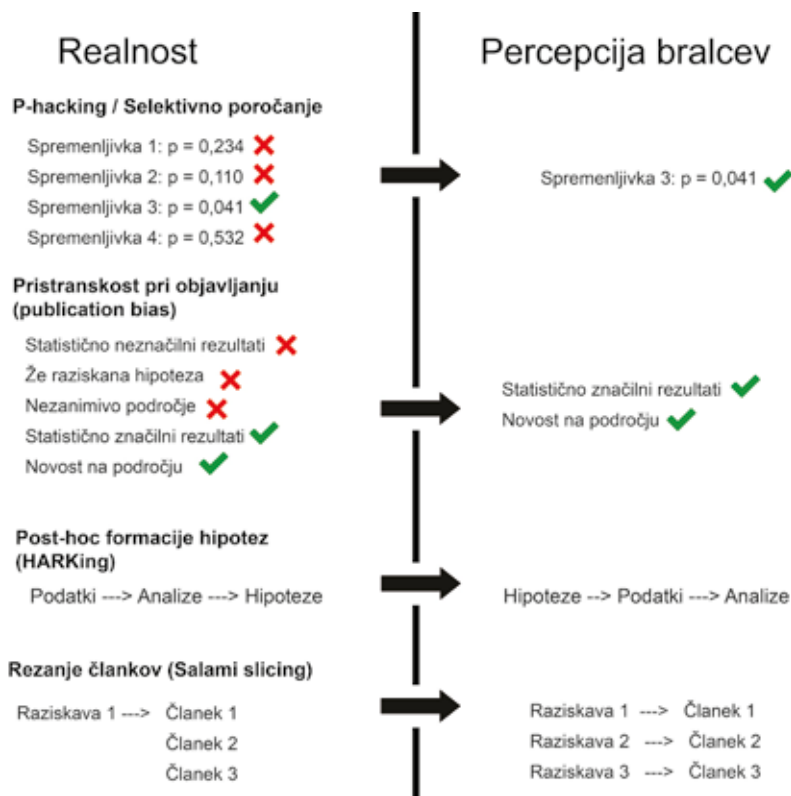
O selektivnem poročanju smo pisali že v poglavju 3.1. Ko raziskovalci (ali revije) objavljajo zgolj izsledke zanimivih raziskav in statistično značilne rezultate, se podoba realnosti izkrivlja. Vsakršno selektivno poročanje o rezultatih je problematično in ni transparentno. Posebna težava je selektivno poročanje o odvisnih spremenljivkah. V raziskavo lahko vključimo več odvisnih spremenljivk, ključno pa je, da se za vse tudi navedejo rezultati oziroma statistične značilnosti (Caldwell idr., 2020). V nasprotnem primeru rezultati raziskave ponovno kažejo izkrivljeno sliko. Vzemimo primer

raziskave, v kateri preučujemo vpliv vadbe proti uporabi na različne krvne markerje. Če raziskovalec pri analizi ugotovi statistično značilen vpliv na dva od 20 preučevanih markerjev in v članku poroča zgolj o statistično značilnih rezultatih, dobi bralec povsem drugačno sliko, kot če so navedeni rezultati za vseh 20 markerjev. Poleg napačnega vtisa gre pri takem selektivnem poročanju tudi za slabo prakso z vidika statistike (glej poglavje 4.6).

P-hacking se širše nanaša na prakso manipuliranja z analizo podatkov za doseganje statistično pomembnih rezultatov (Caldwell idr., 2020; Silberzahn idr., 2018). Ta manipulacija lahko poteka na različne načine, med njimi je tudi prej opisano selektivno poročanje o ugotovitvah. P-hacking vključuje tudi testiranje velikega števila hipotez in iskanje na slepo (izvajanje eksplorativne analize z velikim številom spremenljivk in testov, brez predhodne hipoteze, da bi našli kakršen koli pomemben rezultat). Najverjetneje redkejša praksa je manipulacija z velikostjo vzorca; lahko gre za povečevanje velikosti vzorca, dokler se ne doseže statistično značilen rezultat, ali pa za predčasno zaključevanje zbiranja podatkov, ko je ta dosežen. Tako kot nekatere prej omenjene prakse je p-hacking povezan z večjo verjetnostjo napačnih (lažno pozitivnih) ugotovitev.

Velikost vzorca v raziskavi

Pri statističnih analizah se lahko pojavijo napake oziroma napačno sklepanje, saj so velikosti vzorca preiskovancev omejene, merski postopki pa niso vedno povsem zanesljivi. Statistična napaka 1. vrste odraža sprejetje alternativne hipoteze (potrditev razlik, razmerij, učinkov), čeprav ta dejansko ni resnična (torej v resnici ni razlik, razmerij, učinkov). Verjetnost napake tipa 1 odraža α -vrednost. Statistična napaka 2. vrste pomeni zavrnitev alternativne hipoteze, čeprav razlike, razmerja ali učinki dejansko obstajajo. Napaka 2. vrste se pogosto pojavlja pri majhnih vzorcih (nimamo dovolj preiskovancev, da bi potrdili razlike). Vrednost napake 2. vrste odraža β -vrednost, večkrat pa poročamo o statistični moči. To izračunamo kot $1 - \beta$. Minimalna želena statistična moč (verjetnost, da ne bo prišlo do napake tipa 2) je 80 %. Da se statističnim napakam izognemo, je treba pred začetkom raziskave izračunati zadostno velikost vzorca. Poleg želene statistične moči in α -vrednosti je za izračun velikosti vzorca treba navesti pričakovano velikost



Slika 1.

Primeri slabih praks pri poročanju in interpretaciji rezultatov raziskav. Delno prirejeno po Caldwell idr. (2020).

učinka (velikost razlike, moč korelacije ipd.). Izračun minimalne velikosti vzorca za študijo zagotavlja ustrezno statistično moč za odkrivanje učinkov, razlik ali povezav, kadar te obstajajo. Premajhen vzorec bo privedel do nizke statistične moči, to pa bo vodilo do neveljavnih in neponovljivih rezultatov (Vankov idr., 2014), medtem ko bo prevelik vzorec povezan z nepotrebnimi stroški, poleg tega pa je tudi etično vprašljiv zaradi nepotrebnih tveganj ali nevšečnosti, ki jih nalaga udeležencem. Analiza 120 naključno izbranih člankov, objavljenih v *Journal of Sports Sciences*, je pokazala, da jih je le 11 % poročalo o vnaprejšnji oceni velikosti vzorca (Abt idr., 2020). V preglednem članku so ocenili, da so bile velikosti vzorcev ustrezno utemeljene le pri 19–35 % študij, objavljenih v revijah s področja kineziologije po vsem svetu (Twomey idr., 2021).

Čeprav se zahteva po utemeljitvi velikosti vzorca vse bolj uveljavlja (*Journal of Sports Sciences* denimo zavrača članke brez recenzentskega postopka, če izračuna velikosti vzorca ni), je treba opozoriti, da njegov izračun ni vedno natančno ali dovolj

pojasnjen. Lahko se pojavi neskladje med statističnim testom, navedenim v izračunu moči (npr. t-test), in primarno analizo, izvedeno v študiji (npr. interakcija skupine s časom iz analize varianc). Večkrat je zaznati sklicevanje na neustrezno velikost učinka (iz predhodne raziskave vzamemo velikost učinka za primerjave med preiskovanci, v raziskavi pa načrtujemo primerjavo med skupinami) ali pa se avtorji zanašajo na pilotne podatke (glej tudi zadnji odstavek poglavja 4.3). Raziskovalci pogosto ne upoštevajo predvidenega osipa preiskovancev pri dolgoročnih raziskavah, prepogosto pa tudi ne navedejo dovolj informacij, da bi bralcem omogočili ponovitev izračunov (Chan idr., 2008; Charles idr., 2009). V sodobni literaturi na področju športa so žal še vedno razširjene majhne velikosti vzorcev in posledično nizka statistična moč. Nedavna metaanaliza o fizioloških učinkih visokointenzivnega intervalnega treninga je denimo vključevala 48 študij s povprečno velikostjo vzorca 15 oseb na skupino (Maturana idr., 2020). Skupni učinek v metaanalizi je bil $d = 0,4$, kar pomeni, da je statistična moč večine vključenih študij manjša od

20 %. Zaradi narave področja je pridobiti velik vzorec velikokrat težavno ali celo nemogoče. Po drugi strani pa je raziskovanje hipotez z majhnimi vzorci (posebej kadar so tudi velikosti učinkov nizke) prav tako težavno in vodi do nezanesljivih rezultatov. Zato raziskovalce spodbujamo in pozivamo, naj dosledno in natančno izračunajo velikosti vzorcev za svoje raziskave, pri čemer naj poiščejo pomoč metodologov oziroma statistikov, če niso prepričani o pravilnosti svojega izračuna.

■ Napake in slabe prakse pri statistični obdelavi podatkov

V biomedicinskih znanostih napake pri statistični obdelavi niso redkost (George idr., 2016; Strasak idr., 2007). Na področju športa in kineziologije deluje malo strokovnjakov za statistiko (Sainani idr., 2021), zato so pri statistični obdelavi podatkov raziskovalci večinoma prepuščeni sami sebi. Temeljna napaka pri statistični obdelavi je uporaba napačne analize. Pri osnovnih analizah je ta napaka morda redkejša, a se kljub temu dogaja, da se za analizo ponovljivosti neustrezno uporabi Pearsonov korelacijski koeficient (Koo in Li, 2016) ali pa namesto analize variance opravi več t-testov. Sainani idr. (2021) svarijo pred uporabo nepreverjenih statističnih metod in opozarjajo na neprimerno modifikacijo metod, kot je analiza glavnih komponent (angl. principal component analysis).

Ena od rešitev problema statističnih napak je vsekakor okrepitev sodelovanja s strokovnjaki za statistiko, na kar je v preteklosti opozarjalo že veliko avtorjev (Casals in Finch, 2018; Nielsen idr., 2020; Sainani idr., 2021). V nadaljevanju povzemamo zgolj nekatere pogoste napake in slabe prakse, ki jih raziskovalci lahko odpravijo sami; pri kompleksnejših analizah pa bo za zmanjševanje pogostosti napak nujno sodelovanje s statistiki.

Napačno poročanje o merah variabilnosti in zaupanja

Večina avtorjev raziskav pri poročanju o rezultatih dosledno vključuje mere razpršenosti, kot sta standardni odklon in kvartilni razmik (tudi interkvartilni razpon, angl. interquartile range). Njihovo vključevanje je pomembno, saj na podlagi mer centralne tendence (povprečje ali mediana) ne moremo vedeti, kako razpršene so vrednosti

posameznikov. Dva nabora podatkov z enakim povprečjem imata lahko zelo različne distribucije oziroma variabilnost, to pa je za celovito interpretacijo rezultatov ključnega pomena. Vendarle pa se pri poročanju o razpršenosti pojavljajo nepravilnosti in nedoslednosti, zaznani je predvsem izmenično uporabo standardnega odklona s standardno napako povprečja (angl. standard error of mean; SEM) in intervali zaupanja (angl. confidence intervals), a omenjeni nista meri razpršenosti, temveč meri zaupanja (Earnest idr., 2018). SEM prikazuje, kako natančno podatki vzorca odražajo prvo povprečje populacije. Izračuna se kot razmerje med standardnim odklonom in kvadratnim korenom velikosti vzorca. Z uporabo SEM lahko izračunamo intervale zaupanja – največkrat se izračunavajo 95-odstotni intervali zaupanja. Dobimo jih tako, da SEM pomnožimo z 1,96. Tako dobimo območje okrog povprečja, za katero smo 95-odstotno gotovi, da vsebuje pravo povprečje populacije. SEM in intervali zaupanja se povečujejo, kadar je razpršenost podatkov večja ali velikost vzorca manjša. V Tabeli 1 so prikazani trije primeri hipotetičnih podatkov o telesni višini. V prvem primeru je razpršenost med preiskovanci velika (standardni odklon = 18 cm), velikost vzorca pa razmeroma nizka ($n = 30$). Posledično je visoka tudi SEM (3,29 cm), intervali zaupanja pa so široki (od 168,5 do 181,4 cm). To pomeni, da lahko za pravo povprečje populacije s 95-odstotno gotovostjo trdimo, da se giblje med 168,5 in 181,4 cm. V drugem primeru gre za podatke z enakim povprečjem in standardnim odklonom, a je vzorec preiskovancev večji ($n = 100$). Kljub enaki razpršenosti oziroma variabilnosti podatkov (enak standardni odklon) smo o natančnosti povprečja bolj prepričani (intervali zaupanja od 171,5 do 178,5 cm). V zadnjem primeru gre za podatke z manjšo razpršenostjo (standardni odklon = 10 cm), SEM in intervali zaupanja se dodatno zmanjšajo. Kot kažejo primeri, smo lahko kljub razmeroma veliki razpršenosti podatkov med preiskovanci (tj. visok standardni odklon) razmeroma dobro prepričani o re-

prezentativnosti povprečja, če smo meritve izvedli na velikem številu preiskovancev.

Raziskovalce pozivamo, naj bodo pri poročanju o merah razpršenosti in zaupanja previdni. Predvsem pri grafičnih prikazih se prepogosto dogaja, da se namesto standardnega odklona prikaže SEM, saj je ta manjša in na grafih zavzame manj prostora. Predlagamo, da se na grafih prikazuje standardni odklon, v besedilu pa se dodatno lahko navedejo intervali zaupanja, medtem ko vključitev SEM v večini primerov ni bistvena (Hopkins idr., 2009). Pri neparametrični statistiki se za grafično ponazoritev priporoča uporaba »škafle z brki« (angl. boxplot). Škatla z brki je grafična predstavitev razpršenosti podatkov s pomočjo mediane, prvega in tretjega kvartila ter najnižje in najvišje vrednosti podatkov. Posebna težava se lahko pojavi, ko poročane vrednosti drugi raziskovalci napačno uporabijo v metaanalizah. Če se namesto standardnega odklona za izračun velikosti učinka uporabi SEM, bo učinek študije v metaanalizi precenjen. Nedavno objavljen pregled literature ugotavlja skrb zbujajočo visoko prevalenco napak v metaanalizah na področju športne znanosti; pri kar 85 % visokocitiranih metaanalizah so odkrili vsaj eno statistično nepravilnost, pri čemer je bila skoraj polovica napak (45 %) povezanih z napačnim izračunom velikosti učinkov zaradi zamenjave standardnega odklona in SEM (Kadlec idr., 2023).

Naj dodamo še, da se zaželenost poročanja o intervalih zaupanja ne nanaša samo na povprečne vrednosti, temveč tudi na druge statistične vrednosti, kot so povprečne razlike (spremembe), velikosti učinkov, korelacijski koeficienti, razmerje obetov in tako naprej. V večini primerov nam programi za statistično analizo intervale zaupanja že ponudijo. Za ponazoritev pomembnosti vključevanja intervalov zaupanja vzemimo še intraklasni koeficient korelacije (angl. intra-class correlation coefficient, ICC), ki ga uporabljamo za oceno relativne ponovljivosti podatkov. Ponovljivost po ICC se interpretira kot slaba ($< 0,50$), zmerna ($0,50-0,75$), dobra ($0,75-0,90$) in odlična

($> 0,90$) (Koo in Li, 2016). Prepogosto se rezultate ponovljivosti glede na ICC oceni samo na podlagi sredinske vrednosti ICC, brez ozira na intervale zaupanja. Vzemimo za primer, da dobimo vrednost ICC = 0,80 s 95-odstotnimi intervali zaupanja od 0,65 do 0,95. Tak rezultat raziskovalci pogosto označijo kot »dobro ponovljivost«. Pravilno in transparentno poročanje bi bilo (skupaj z intervali zaupanja v besedilu ali tabeli), da je ponovljivost »zmerna do odlična« (Giuseppe, 2018; Koo in Li, 2016).

Zanašanje zgolj na p-vrednosti (neporočanje o velikosti učinkov)

Pri zajemanju podatkov se vedno pojavljajo napake in naključna variacija. V športni znanosti in kineziologiji, tako kot pri številnih drugih vedah, statistično značilne rezultate (razlike med povprečji, korelacijske koeficiente ipd.) sprejemamo pri vrednosti $p < 0,05$ (Tiller in Ekkekakis, 2023), s tem pa je povezanih veliko napak in slabih praks. Izhajali bomo iz testiranja razlik, a opisano velja za vse statistične teste. Začnemo lahko pri razumevanju te vrednosti: p-vrednost se pogosto napačno predstavlja kot verjetnost, da so ugotovljene razlike posledica naključja (vrednost $p = 0,05$ naj bi pomenila, da je verjetnost, da so ugotovljene razlike posledica naključja, le 5 %; z drugimi besedami, statistično značilen rezultat potrdimo, ko je manj kot 5 % verjetnosti, da so razlike posledica naključja). Vendar je prava interpretacija p-vrednosti nekoliko drugačna, in sicer gre za verjetnost pridobitve rezultatov testa (npr. razlik), ki so vsaj tako ekstremni kot dejansko opaženi rezultat, ob začetni predpostavki, da je ničelna hipoteza (da razlik ni) pravilna (Andrade, 2019). Vzemimo za primer, da izmerimo razliko v jakosti stiska pesti med nogometaši in rokometaši z vrednostjo $p = 0,03$. Napačno bi bilo trditi, da je zgolj 3 % možnosti, da bi take razlike dobili naključno oziroma da smo 97 % gotovi, da so rezultati odraz dejanskih razlik. Obratno, ta p-vrednost pove, da je zgolj 3 % možnosti, da bi dobili take razlike, če v resnici razlik ni

Tabela 1.

Prikaz odnosa med povprečjem, razpršenostjo podatkov, standardno napako povprečja in intervali zaupanja.

Primer	P	SO	n	SEM	95% intervali zaupanja
1	175 cm	18 cm	30	$18 / \sqrt{30} = \mathbf{3,29 \text{ cm}}$	$175 \pm (1,96 \times 3,29) = \mathbf{168,5 - 181,4 \text{ cm}}$
2	175 cm	18 cm	100	$18 / \sqrt{100} = \mathbf{1,8 \text{ cm}}$	$175 \pm (1,96 \times 1,8) = \mathbf{171,4 - 178,5 \text{ cm}}$
3	175 cm	10 cm	100	$10 / \sqrt{100} = \mathbf{1 \text{ cm}}$	$175 \pm (1,96 \times 1) = \mathbf{173,0 - 176,9 \text{ cm}}$

P – povprečje; SO – standardni odklon; n – velikost vzorca; SEM – standardna napaka povprečja

(ničelna hipoteza). Drugače povedano, če bi pri nogometaših in rokometiših izmerili popolnoma enake povprečne vrednosti, bi tako velike razlike dobili le 3-krat, če bi raziskavo 100-krat ponovili. Pri opisani (napačni) razlagi je p -vrednost nepravilno vzeta kot neposredna indikacija verjetnosti za resničnost razlik. Pri pravilni razlagi p -vrednost predstavlja možnosti skrajnosti opaženih podatkov pod predpostavko, da je ničelna hipoteza resnična (da razlik ni), ne obravnava neposredno resničnosti oziroma verjetnosti same hipoteze.

Druga problematika je binarnost odločanja na podlagi p -vrednosti (rezultat je lahko le statistično značilen ali ne), a to ne kaže celotne slike. Zato p -vrednosti ne bi smeli interpretirati izolirano, temveč jo je treba obravnavati skupaj z drugimi statističnimi vrednostmi, ki jim lahko dodamo intervale zaupanja. Pri testiranju razlik so to mere velikosti učinka (angl. Effect size), kot so Cohenov d in eta-kvadrat (Bakeman, 2005; Cohen, 1988). Dodatno je smiselno poročati o povprečnih razlikah, ki jim prav tako lahko izračunamo z intervale zaupanja. Velikost učinka je kvantitativno merilo velikosti pojava (razlik, povezav itd.), uporablja pa se za interpretacijo praktičnega in kliničnega pomena ugotovitev raziskave, ki presega zgolj statistično pomembnost, označeno s p -vrednostmi (Bakker idr., 2019; Tiller in Ekkekakis, 2023). Nekateri avtorji celo menijo, da so velikosti učinkov najpomembnejši podatek pri poročanju o rezultatih eksperimentalnih študij (Lakens, 2013). Analize literature kažejo, da le slaba tretjina raziskav s področja športne prehrane dosledno poroča o velikosti učinka (Earnest idr., 2018).

Kot smo že omenili, je zaželeno, da se velikostim učinka pripnejo intervale zaupanja (Schulz idr., 2010). To je to ključnega pomena za interpretacijo zaupanja v velikost učinka. Vzemimo za primer, da raziskujemo vpliv vadbe proti uporu na mišično togost. Po intervenciji zabeležimo statistično značilne razlike (npr. pri vrednosti $p = 0,011$). Dodajanje velikosti učinka k poročanju je nujna, da vidimo, kako velik je bil ta učinek (ob velikem vzorcu preiskovancev so lahko tudi zelo majhni učinki statistično značilni). Če na primer dobimo (po Cohenovem d) velikost učinka $d = 0,95$, poročamo o visokem učinku vadbe proti uporu na mišično togost. V tem primeru se interpretacija ubogati, saj vidimo, da ima vadba velike učinke. Vendar moramo za popolnost poročanja vključiti tudi intervale zaupanja za velikost učinka (Bakker idr., 2019); če deni-

mo dobimo široke intervale zaupanja, na primer d -vrednost od 0,27 do 1,45, potem smo lahko 95-odstotno gotovi le, da je resnični učinek vadbe v okviru teh vrednosti (in je torej lahko tako majhen ali srednji kot tudi visok).

Posebna težava se pojavi, ko velikosti učinka (iz predhodnih raziskav ali na podlagi lastnih pilotnih meritev) uporabljamo za izračun potrebne velikosti vzorca za prihodnjo raziskavo (glej poglavje 3.5). Vzemimo za primer vrednost velikosti učinka iz prejšnjega odstavka ($d = 0,95$). Izračun potrebne velikosti vzorca za preprost parni t -test pri 90-odstotni statistični moči in α -vrednosti 0,05 kaže, da za raziskavo potrebujemo zgolj 11 preiskovancev. Ob upoštevanju spodnje meje intervala zaupanja za velikost učinka ($d = 0,27$) pa program navede, da potrebujemo kar 119 preiskovancev. Gre sicer za nekoliko skrajni primer, ki nam je v pomoč za ponazoritev, a podobne napake se vsekakor pojavljajo. Raziskovalcem zato svetujemo, naj pri poročanju o velikosti učinkov navedejo 95-odstotne intervale zaupanja tudi za bolj natančno oceno velikosti vzorcev v prihodnjih raziskavah. Posebno izrazite napake se dogajajo, ko velikost učinka določimo na podlagi pilotne raziskave z majhnim številom preiskovancev; ob izračunu intervalov zaupanja bi videli, da je natančnost ocene predvidene velikosti učinka majhna.

Neupoštevanje kovariat

Pogosta slaba praksa je tudi neupoštevanje spremenljivk, ki lahko vplivajo na rezultate osnovne analize. Sainani idr. (2021) kot dober primer navajajo raziskavo o povezavi med koncentracijo vitamina D v krvi in verjetnostjo za pojav menstrualnih težav pri mlajših ženskah (Łagowska, 2018). V raziskavi so poročali, da je koncentracija vitamina D v krvi pod 30 ng/mL povezana s petkrat večjo verjetnostjo za pojav menstrualnih težav. Natančnejši pregled rezultatov pa razkriva potencialen vpliv kovariate, ki je raziskovalci niso upoštevali. Medtem ko je večji delež žensk z nizko koncentracijo vitamina D (40 % od skupno 60 preiskovank) imel menstrualne težave v primerjavi s skupino z visoko koncentracijo vitamina D (12 % od skupno 17), analiza ni upoštevala razlik v telesni masi. Skupina z nizko vsebnostjo vitamina D je imela za 17 % višjo telesno maso od skupine z visoko koncentracijo vitamina D (povprečna telesna masa 66,7 kg proti 57,0 kg). Obenem je bila tudi telesna masa močno povezana s

pojavnostjo menstrualnih težav (preiskovanke s težavami so imele povprečno telesno maso 77,6 kg, ženske brez težav pa 57,9 kg). Torej je navidezno razmerje med nizkim vitaminom D in tveganjem za menstrualne težave deloma ali v celoti posledica vpliva telesne mase na menstrualne težave. Upoštevanje kovariat ni le statistična praksa, temveč izhaja iz strokovne presoje. Raziskovalce spodbujamo, naj pri snovanju načrta raziskav razmislijo, katere kovariate bi bilo treba upoštevati, da bo primarno raziskovalno vprašanje preučeno z večjo mero veljavnosti. V raziskavah učinkov vadbe je kot kovariato posebej pomembno upoštevati začetno raven treniranosti pri preiskovancih, saj so lahko od te močno odvisni odzivi na vadbeno intervencijo. Seveda je zaželeno, da z velikim vzorcem preiskovancev in randomizacijo potencialne razlike v stopnji začetne treniranosti med vadbeno in kontrolno skupino preiskovancev minimiziramo (Hecksteden idr., 2018).

Nevključevanje statističnih korekcij

Statistične korekcije so pri večkratnih analizah pomembne zato, da se izognemo statistični napaki tipa 1. Drugače povedano, ko se hkrati testira več hipotez ali spremenljivk, se verjetnost potrditve vsaj ene hipoteze poveča z vsakim dodatnim testom. Zato se za ohranitev skupne stopnje napake na sprejemljivi ravni zahteva uporaba statističnih korekcij (Altman in Bland, 1995). Raziskovalcem bolj znane so korekcije pri parnih primerjavah več skupin ali setov podatkov v različnih časovnih točkah, kar je običajno opravljeno po analizi variance. Denimo, da želimo oceniti razlike med tremi skupinami športov pri časih sprinta na 100 metrov. Izvede se enosmerna analiza variance, da se ugotovi, ali obstajajo statistično značilne razlike med tremi skupinami (tj. da se vsaj ena skupina statistično značilno razlikuje od drugih dveh). Da bi ugotovili, katere specifične skupine se med seboj razlikujejo, so potrebne post hoc primerjave. Na podlagi več opravljenih primerjav (skupina A proti B, skupina A proti C, skupina B proti C) se za kontrolo napak tipa 1 največkrat uporabi Bonferronijev popravek. S tremi primerjavami je α -vrednost prilagojena tako, da jo delimo s številom primerjav ($0,05/3 = 0,0167$). Denimo, da post hoc testi kažejo naslednje: skupina A proti skupini B ($p = 0,02$); skupina A proti skupini C ($p = 0,01$) in skupina B proti skupini C ($p = 0,015$). Glede na prilagojeno α -vrednost $= 0,0167$ so statistično značilne samo razlike

med skupino A in skupino C ($p = 0,01$) ter skupino B in skupino C ($p = 0,015$). Čeprav primerjava med skupino A in skupino B kaže p -vrednost 0,02, se po Bonferronijevi prilagoditvi ne šteje za pomembno. S prilagajanjem praga pomembnosti analiza zagotavlja, da ugotovljene pomembne razlike med režimi vadbe niso naključne.

Popravki pa niso pomembni le v primeru več parnih primerjav. Če izvedemo dva neodvisna testa za dve ločeni spremenljivki, se verjetnost napake tipa 1 poveča s 5 % na 10 %, pri analizi šestih spremenljivk na 26 % ter pri 14 spremenljivkah na 50 % (Tiller in Ekkekakis, 2023). Namerno izogibanje korekcijam se zato šteje za enega od načinov »p-hackinga«. Povedano preprosto, z vključevanjem velikega števila odvisnih spremenljivk povečamo možnost, da bo vsaj ena analiza statistično značilna. Analiza literature na področju športa kaže, da je vrednost mediane izvedenih statističnih testov na posamezen članek kar 30, ob tem pa je le pri 14 % člankov jasno navedeno, katera odvisna spremenljivka je primarna (Lohse idr., 2020; Sainani in Chamari, 2022). Raziskovalce spodbujamo, naj pri analizi več odvisnih spremenljivk vključijo primereno korekcijo. Dilema pri korekcijah v tem primeru se pojavi, ker zniževanje vrednosti α avtomatsko poveča napako tipa 2, saj bomo hipoteze težje potrdili, tudi če so resnične. Zato se namesto Bonferronijevega popravka v teh primerih uporabljajo manj konservativni pristopi. Primer take korekcije je Holm-Bonferronijeva metoda (Chen idr., 2017). Pri tej se najnižja p -vrednost primerja s prilagojeno α -vrednostjo po klasični Bonferronijevi metodi (α deljena s številom odvisnih spremenljivk), vsaka naslednja p -vrednost pa z nekoliko večjo alfo.

Neupoštevanje porazdelitve podatkov

Dobra praksa, ki jo priporočamo vsem raziskovalcem, je vizualizacija podatkov pred izvedbo analiz. Za uporabo parametrične statistike (t-testi, analize variance, Pearsonov korelacijski koeficient ipd.) je zahtevana normalna porazdelitev podatkov (tudi Gaussova porazdelitev). Ta se v grafični predstavitvi kaže v obliki simetričnega zvona z največjo frekvenco vrednosti ob povprečju (Yap in Sim, 2011). Za vizualno preverbo normalnosti porazdelitve podatkov se svetuje uporaba grafikona kvantilov (angl. quantile-quantile plot) (Loy idr., 2016). Poleg vizualne preverbe se zahteva tudi statistično preverjanje normalnosti

porazdelitve. To se največkrat opravi s Shapiro-Wilkovim testom, tudi rezultate tega je smiselno navesti v članku. Ob morebitnem odstopanju od normalnosti porazdelitve se za nadaljnjo analizo uporabijo neparametrični testi. Pri tem se pojavlja skrb pred večjim številom napak, saj strokovnjaki na področju športa svoje znanje o neparametričnih testih ocenjujejo kot precej slabše v primerjavi z znanjem o parametričnih testih (Ocakoglu idr., 2020).

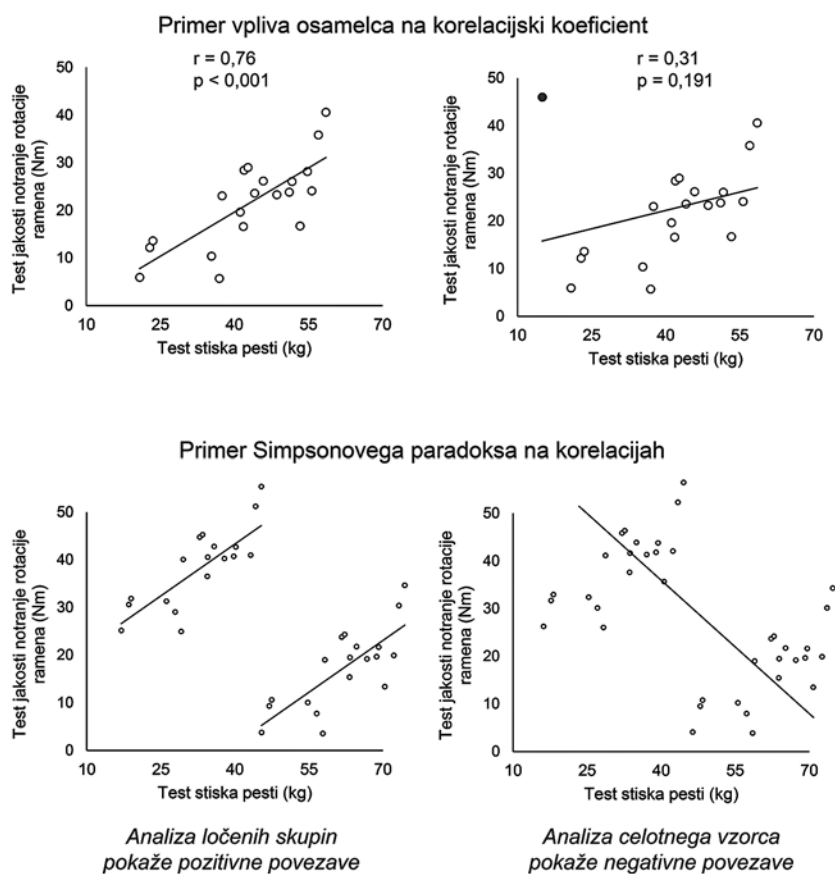
Normalnost porazdelitve ni edina anomalija v podatkih, ki lahko pomembno vpliva na rezultate analize. Vizualna preverba podatkov lahko pokaže osamelce (podatke, ki izrazito odstopajo od preostalih), ki so lahko posledica napake v meritvah ali izračunih. Izrazit osamelec lahko močno vpliva na povprečno vrednost in s tem na rezultate analiz. Prav tako lahko močno popači korelacijske koeficiente (Slika 2, zgoraj). Vizualizacijo podatkov je smiselno opraviti tudi ločeno po skupinah (npr. po spolu, športu ipd.). V redkih primerih lahko pride do porazdelitve podatkov, pri kateri je korelacija med spremenljivkama v posameznih

skupinah obratna kot na celotnem vzorcu (Slika 2, spodaj). Gre za enega od tipov Simpsonovega paradoksa.

Poziv k izboljšanju

Ključna do sprememb sta po našem mnenju dva: prvi je, da se raziskovalci zavedamo potencialnih napak in se jim proaktivno izogibamo, drugi pa je okrepitev sodelovanja s statistiki. Trenutno naj bi le okoli 13 % člankov na področju športne znanosti vključevalo vsaj enega avtorja s področja statistike ali sorodnega področja (podatkovna znanost, epidemiologija) (Sainani idr., 2021). Ukrepi, ki so kratko opisani v tem poglavju, so namenjeni predvsem večji transparentnosti, to pa posredno pripomore k odpravi nekaterih slabih praks, kot so t. i. »p-hacking«, »HARKing« in »salami slicing«.

V zadnjem času se v luči spodbujanja odprte znanosti vse bolj poudarja tudi dostopnost oziroma razpoložljivost surovih podatkov, pridobljenih v raziskavi (McGuin-



Slika 2. Vizualizacija podatkov lahko pomembno vpliva na interpretacijo statističnih analiz (primer korelacij).

ness in Sheppard, 2021). Razpoložljivost podatkov je ključnega pomena za zagotavljanje transparentnosti in ponovljivosti, dveh pomembnih dejavnikov za ohranjanje integritete znanstvenih raziskav. Ko so podatki zlahka dostopni, lahko drugi raziskovalci preverijo ustreznost statističnih analiz in s tem ugotovitev raziskave. Razpoložljivost podatkov prav tako krepi zaupanje v znanstvene rezultate v znanstveni skupnosti in javnosti ter spodbuja sodelovanje. Vse več revij ob oddaji članka zahteva izjavo o razpoložljivosti podatkov (angl. Data availability statement). Tudi v športni znanosti se vse bolj prepoznava potencial deljenja in združevanja podatkov v večje baze (Passfield in Hopker, 2017). Podatki se lahko delijo na več načinov, med drugim v obliki dodatnega gradiva (angl. Supplementary materials) pri oddaji članka. Podatke lahko naložimo na različne namenske spletne portale, kot so Zenodo, Open Science Framework in podobno. Prav tako je mogoče podatke naložiti na portal ResearchGate. Pri tem naj bodo avtorji pozorni na ustrezno anonimizacijo podatkov.

Dodatno transparentnost raziskovanja lahko dosežemo z naprejšnjo registracijo raziskav. Popis hipotez in primernih odvisnih spremenljivk pred raziskavo prepreči p-hacking, HARKing in selektivno poročanje. Eksperimentalne klinične študije se tipično registrirajo na portalih, kot je ClinicalTrials.gov (nekateri revije, tudi s področja športne znanosti, to registracijo zahtevajo) (Chahal idr., 2012). Vse več je možnosti, da raziskovalci v namenski reviji oddajo v recenzijo protokol raziskave in tako prejmejo povratno informacijo še preden izvedbo raziskave. Posebna vrsta publikacije, ki počasi pridobiva svoje mesto tudi v znanosti o športu, je »Registrirano poročilo« (angl. Registered reports) (Caldwell idr., 2020; Hardwicke in Ioannidis, 2018). Pri tej obliki objavljanja avtorji v recenzijo oddajo protokol raziskave. Če je ta ustrezen, se objava preliminarno sprejme; če avtorji nato raziskavo izvedejo skladno s protokolom, je končni članek sprejet ne glede na rezultate oziroma statistično značilnost. Opažamo tudi porast uporabe portalov, kamor lahko avtorji oddajo prednatis (angl. preprint) članka, preden je ta recenziran in sprejet v objavo. Poleg hitrejši diseminacije rezultatov lahko to pripomore k širši povratni informaciji (in s tem možnosti za popravke) pred končno objavo članka. Številne revije že ponujajo možnost, da se ob oddaji članka v recenzijo ta avtomatsko naloži na portal za prednatis. Za področje športne

znanosti in kineziologije je trenutno najbolj uveljavljen portal SportRxiv (Caldwell, 2023). Raziskovalce torej spodbujamo, da na tak ali drugačen način vnaprej »registrirajo« svojo raziskavo ter s tem poskrbijo za transparentnost in integriteto končnih znanstvenih objav. Zavedamo se, da je pri tem nekaj omejitev, saj vse to zahteva dodatno delo, hkrati pa se lahko nekoliko podaljša tudi čas do objave.

Nazadnje bi priporočili še, da se raziskovalci tako pri snovanju eksperimentov kot pri poročanju o rezultatih opirajo na relevantne smernice: CONSORT pri eksperimentalnih randomiziranih raziskavah (Schulz idr., 2010), STROBE pri opazovalnih študijah (Von Elm idr., 2007), GRRAS za študije ponovljivosti (Kottner idr., 2011) in smernice PRISMA pri pripravi sistematičnih pregledov z metaanalizo (Page idr., 2021).

■ Zaključek

Na podlagi predhodnih pregledov in opozoril v tuji literaturi smo pripravili pregled metodoloških pomanjkljivosti in izzivov, povezanih z znanstveno integriteto v raziskavah znanosti o športu. Poleg ozaveščanja o najpogostejših slabih praksah in okrepitvi sodelovanja s statistiki je ključno, da raziskovalci aktivno delujemo v smeri povečanja transparentnosti v raziskovalnem procesu. Posebej poudarjamo pomen dostopnosti surovih podatkov in vnaprejšnje registracije protokolov raziskav, saj se s tem lahko prepreči katera od obravnavanih slabih praks. V skladu z načeli odprte znanosti in etičnimi smernicami se vse bolj poudarja pomen kulture sodelovanja in transparentnega raziskovalnega procesa. Z zavedanjem odgovornosti do znanstvene skupnosti in družbe lahko prispevamo k bolj zanesljivim in uporabnim raziskavam za napredek na področju športa in kineziologije. Naj še enkrat poudarimo, da smo v članku obravnavali le nekatere izmed najpogostejših napak in praks, ki jih omenjajo tudi tuji avtorji. Članek tako ne vključuje popisa vseh možnih napak in slabih praks.

■ Literatura

1. Abt, G., Boreham, C., Davison, G., Jackson, R., Nevill, A., Wallace, E. in Williams, M. (2020). Power, precision, and sample size estimation in sport and exercise science research. *Journal of Sports Sciences*, 38(17), 1933–1935. <https://doi.org/10.1080/02640414.2020.1776002>

2. Abt, G., Jobson, S., Morin, J. B., Passfield, L., Sampaio, J., Sunderland, C. in Twist, C. (2022). Raising the bar in sports performance research. *Journal of Sports Sciences*, 40(2), 125–129. <https://doi.org/10.1080/02640414.2021.2024334>
3. Aksnes, D. W., Langfeldt, L. in Wouters, P. (2019). Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *SAGE Open*, 9(1). <https://doi.org/10.1177/2158244019829575>
4. Altman, D. G. in Bland, J. M. (1995). Multiple significance tests: the Bonferroni method. *BMJ*, 310(6973), 170.
5. Andrade, C. (2019). The P value and statistical significance: Misunderstandings, explanations, challenges, and alternatives. *Indian Journal of Psychological Medicine*, 41(3), 210–215. https://doi.org/10.4103/IJPSYM.IJP-SYM_193_19
6. Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3), 379–384. <https://doi.org/10.3758/BF03192707>
7. Bakker, A., Cai, J., English, L., Kaiser, G., Mesa, V. in Van Dooren, W. (2019). Beyond small, medium, or large: points of consideration when interpreting effect sizes. *Educational Studies in Mathematics*, 102(1). <https://doi.org/10.1007/s10649-019-09908-4>
8. Begley, C. G. in Ioannidis, J. P. A. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Research*, 116(1), 116–126. <https://doi.org/10.1161/CIRCRESAHA.114.303819>
9. Bernards, J. R., Sato, K., Haff, G. G. in Bazylar, C. D. (2017). Current research and statistical practices in sport science and a need for change. *Sports*, 5(4). <https://doi.org/10.3390/sports5040087>
10. Borg, D. N., Barnett, A. G., Caldwell, A. R., White, N. M. in Stewart, I. B. (2023). The bias for statistical significance in sport and exercise medicine. *Journal of Science and Medicine in Sport*, 26(3), 164–168. <https://doi.org/10.1016/j.jsams.2023.03.002>
11. Büttner, F., Toomey, E., McClean, S., Roe, M. in Delahunt, E. (2020). Are questionable research practices facilitating new discoveries in sport and exercise medicine? The proportion of supported hypotheses is implausibly high. *British Journal of Sports Medicine*. <https://doi.org/10.1136/bjsports-2019-101863>
12. Caldwell, A. R. (2023). *A Tutorial on How to Utilize SportRxiv: Submission Tutorial*. SportRxiv. <https://doi.org/https://doi.org/10.51224/SRXIV.293>
13. Caldwell, A. R., Vigotsky, A. D., Tenan, M. S., Radel, R., Mellor, D. T., Kreutzer, A., Lahart, I. M., Mills, J. P. in Boisgontier, M. P. (2020). Moving Sport and Exercise Science Forward: A Call for the Adoption of More Transparent Research Practices. *Sports Medicine*, 50(3),

- 449–459. <https://doi.org/10.1007/s40279-019-01227-1>
14. Callaway, E. (2011). Fraud investigation rocks Danish university. *Nature*, 7. <https://doi.org/10.1038/news.2011.703>
15. Casals, M. in Finch, C. F. (2018). Sports Biostatistician: A critical member of all sports science and medicine teams for injury prevention. *British Journal of Sports Medicine*, 52(22), 1457–1461. <https://doi.org/10.1136/bjsports-2016-042211rep>
16. Chahal, J., Tomescu, S. S., Ravi, B., Bach, B. R., Ogilvie-Harris, D., Mohamed, N. N. in Gandhi, R. (2012). Publication of sports medicine-related randomized controlled trials registered in clinicaltrials.gov. *American Journal of Sports Medicine*, 40(9), 1970–1977. <https://doi.org/10.1177/0363546512448363>
17. Chan, A. W., Hróbjartsson, A., Jørgensen, K. J., Gøtzsche, P. C. in Altman, D. G. (2008). Discrepancies in sample size calculations and data analyses reported in randomised trials: Comparison of publications with protocols. *Bmj*, 337(7683), 1404–1407. <https://doi.org/10.1136/bmj.a2299>
18. Charles, P., Giraudeau, B., Dechartres, A., Baron, G. in Ravaud, P. (2009). Reporting of sample size calculation in randomised controlled trials: Review. *BMJ (Online)*, 338(7705), 1256. <https://doi.org/10.1136/bmj.b1732>
19. Chen, S. Y., Feng, Z. in Yi, X. (2017). A general introduction to adjustment for multiple comparisons. *Journal of Thoracic Disease*, 9(6), 1725–1729. <https://doi.org/10.21037/jtd.2017.05.34>
20. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge Academic.
21. Coursol, A. in Wagner, E. E. (1986). Effect of positive findings on submission and acceptance rates: A note on meta-analysis bias. *Professional Psychology: Research and Practice*, 17(2), 136–137. <https://doi.org/10.1037/0735-7028.17.2.136>
22. Dahlberg, J. E. in Mahler, C. C. (2006). The poehlman case: running away from the truth. *Science and Engineering Ethics*, 12(1), 157–173. <https://doi.org/10.1007/s11948-006-0016-9>
23. Earnest, C. P., Roberts, B. M., Harnish, C. R., Kutz, J. L., Cholewa, J. M. in Johannsen, N. M. (2018). Reporting characteristics in sports nutrition. *Sports*, 6(4). <https://doi.org/10.3390/sports6040139>
24. Ekmekci, P. E. (2017). An increasing problem in publication ethics: Publication bias and editors' role in avoiding it. *Medicine, Health Care and Philosophy*, 20(2), 171–178. <https://doi.org/10.1007/s11019-017-9767-0>
25. Gaspar, D. E. P. in Esteves, M. D. L. (2021). Awareness of the Misconduct in Sports Science Research. *Annals of Applied Sport Science*, 9(3), 0–0. <https://doi.org/10.52547/aassjournal.934>
26. George, B. J., Beasley, T. M., Brown, A. W., Dawson, J., Dimova, R., Divers, J., Goldsby, T. U., Heo, M., Kaiser, K. A., Keith, S. W., Kim, M. Y., Li, P., Mehta, T., Oakes, J. M., Skinner, A., Stuart, E. in Allison, D. B. (2016). Common scientific and statistical errors in obesity research. *Obesity*, 24(4), 781–790. <https://doi.org/10.1002/oby.21449>
27. Giuseppe, P. (2018). StaTips Part IV: Selection, interpretation and reporting of the intraclass correlation coefficient. *South European Journal of Orthodontics and Dentofacial Research*, 5(1), 3–5. <https://doi.org/10.5937/sejodr5-17434>
28. Gopalakrishna, G., ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M. in Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in the Netherlands. *PLoS ONE*, 17(2 February). <https://doi.org/10.1371/journal.pone.0263023>
29. Hardwicke, T. E. in Ioannidis, J. P. A. (2018). Mapping the universe of registered reports. *Nature Human Behaviour*, 2(11), 793–796. <https://doi.org/10.1038/s41562-018-0444-y>
30. Hecksteden, A., Faude, O., Meyer, T. in Donath, L. (2018). How to construct, conduct and analyze an exercise training study? *Frontiers in Physiology*, 9(JUL). <https://doi.org/10.3389/fphys.2018.01007>
31. Hopkins, W. G., Marshall, S. W., Batterham, A. in Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise*, 41(1), 3–12. <https://doi.org/10.1249/MSS.0b013e31818cb278>
32. Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
33. John, L. K., Loewenstein, G. in Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, 23(5), 524–532. <https://doi.org/10.1177/0956797611430953>
34. Kadlec, D., Sainani, K. L. in Nimphius, S. (2023). With Great Power Comes Great Responsibility: Common Errors in Meta-Analyses and Meta-Regressions in Strength & Conditioning Research. *Sports Medicine*, 53(2), 313–325. <https://doi.org/10.1007/s40279-022-01766-0>
35. Kardeş, S., Levack, W., Özkuk, K., Atmaca Aydın, E. in Serinç Karabulut, S. (2020). Retractions in Rehabilitation and Sport Sciences Journals: A Systematic Review. *Archives of Physical Medicine and Rehabilitation*, 101(11), 1980–1990. <https://doi.org/10.1016/j.apmr.2020.03.010>
36. Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4
37. Koo, T. K. in Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcmm.2016.02.012>
38. Kottner, J., Audige, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., Roberts, C., Shoukri, M. in Streiner, D. L. (2011). Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *International Journal of Nursing Studies*, 48(6), 661–671. <https://doi.org/10.1016/j.ijnurstu.2011.01.016>
39. Łagowska, K. (2018). The relationship between vitamin d status and the menstrual cycle in young women: A preliminary study. *Nutrients*, 10(11). <https://doi.org/10.3390/nu10111729>
40. Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 1–12.
41. Landhuis, E. (2016). Scientific literature: Information overload. *Nature*, 535(7612), 457–458. <https://doi.org/10.1038/nj7612-457a>
42. Lohse, K. R., Sainani, K. L., Taylor, J. A., Butson, M. L., Knight, E. J. in Vickers, A. J. (2020). Systematic review of the use of „magnitudebased inference“ in sports science and medicine. *PLoS ONE*, 15(6). <https://doi.org/10.1371/journal.pone.0235318>
43. Loy, A., Follett, L. in Hofmann, H. (2016). Variations of Q–Q plots: the power of our eyes! *The American Statistician*, 70(2), 202–214. <https://doi.org/10.1080/00031305.2015.1077728>
44. Maggio, L., Dong, T., Driessen, E. in Artino, A. (2019). Factors associated with scientific misconduct and questionable research practices in health professions education. *Perspectives on Medical Education*, 8(2), 74–82. <https://doi.org/10.1007/s40037-019-0501-x>
45. Maturana, F. M., Martus, P., Zipfel, S. in Niess, A. M. (2020). Effectiveness of HIE versus MICT in improving cardiometabolic risk factors in health and disease: a meta-analysis. *Medicine and Science in Sports and Exercise*, 53(3), 559–573. <https://doi.org/10.1249/MSS.0000000000002506>
46. McGuinness, L. A. in Sheppard, A. L. (2021). A descriptive analysis of the data availability statements accompanying medRxiv preprints and a comparison with their published counterparts. *PLoS ONE*, 16(5 May). <https://doi.org/10.1371/journal.pone.0250887>
47. Nielsen, R. O., Shrier, I., Casals, M., Nettel-Aguirre, A., Møller, M., Bolling, C., Bittencourt, N. F. N., Clarsen, B., Wedderkopp, N., Soligard, T., Timpka, T., Emery, C., Bahr, R., Jacobsson, J., Whiteley, R., Dahlstrom, O., Van Dyk, N., Pluim, B. M., Stamatakis, E., ... Verhagen, E. (2020). Statement on methods in sport injury research from the 1st METHODS MATTER Meeting, Copenhagen, 2019. *British Journal of*

- Sports Medicine*, 54(15), 941–947. <https://doi.org/10.1136/bjsports-2019-101323>
48. Nosek, B. A., Ebersole, C. R., DeHaven, A. C. in Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
49. Ocakoglu, G., Macunluoglu, A. C., Can, F. E., Kaymak, B. in Yilvik, Z. (2020). The opinion of sports science professionals for the benefit of statistics: an international web-based survey. *The European Research Journal*, 6(2), 145–153. <https://doi.org/10.18621/eurj.468686>
50. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLoS Medicine*, 18(3). <https://doi.org/10.1371/JOURNAL.PMED.1003583>
51. Passfield, L. in Hopker, J. G. (2017). A mine of information: Can sports analytics provide wisdom from your data? *International Journal of Sports Physiology and Performance*, 12(7), 851–855. <https://doi.org/10.1123/ijspp.2016-0644>
52. Rawat, S. in Meena, S. (2014). Publish or perish: Where are we heading? *Journal of Research in Medical Sciences*, 19(2), 87–89.
53. Robertson, S. (2020). Man & machine: Adaptive tools for the contemporary performance analyst. *Journal of Sports Sciences*, 38(18), 2118–2126. <https://doi.org/10.1080/02640414.2020.1774143>
54. Sainani, K. L., Borg, D. N., Caldwell, A. R., Butson, M. L., Tenan, M. S., Vickers, A. J., Vigotsky, A. D., Warmenhoven, J., Nguyen, R., Lohse, K. R., Knight, E. J. in Bargary, N. (2021). Call to increase statistical collaboration in sports science, sport and exercise medicine and sports physiotherapy. *British Journal of Sports Medicine*, 55(2), 118–122. <https://doi.org/10.1136/bjsports-2020-102607>
55. Sainani, K. L. in Chamari, K. (2022). Wish List for Improving the Quality of Statistics in Sport Science. *International Journal of Sports Physiology and Performance*, 17(5), 673–674. <https://doi.org/10.1123/ijspp.2022-0023>
56. Sasaki, K. in Tan, S. (2018). Publication ethic (1) "salami slicing". *Journal of Hepato-Biliary-Pancreatic Sciences*, 25(6), 321–321. <https://doi.org/10.1002/jhbp.561>
57. Schulz, K. F., Altman, D. G. in Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology*, 63(8), 834–840. <https://doi.org/10.1016/j.jclinepi.2010.02.005>
58. Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahnik, Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Rosa, A. D., Dam, L., Evans, M. H., Cervantes, I. F., ... Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, 1(3), 337–356. <https://doi.org/10.1177/2515245917747646>
59. Smart, P. (2017). Redundant publication and salami slicing: the significance of splitting data. *Developmental Medicine & Child Neurology*, 59(8), 775. <https://doi.org/10.1111/dmcn.13485>
60. Strasak, A. M., Zaman, Q., Pfeiffer, K. P., Göbel, G. in Ulmer, H. (2007). Statistical errors in medical research - A review of common pitfalls. *Swiss Medical Weekly*, 137(3–4), 44–49. <https://doi.org/10.4414/smw.2007.11587>
61. Tiller, N. B. in Ekkekakis, P. (2023). Overcoming the „Ostrich Effect“: A Narrative Review on the Incentives and Consequences of Questionable Research Practices in Kinesiology. *Kinesiology Review*, 12(3), 201–216. <https://doi.org/10.1123/kr.2022-0039>
62. Twomey, R., Yingling, V., Warne, J., Schneider, C., McCrum, C., Atkins, W., Murphy, J., Romero Medina, C., Harlley, S. in Caldwell, A. (2021). Nature of Our Literature. *Communications in Kinesiology*, 1(3). <https://doi.org/10.51224/cik.v1i3.43>
63. Vankov, I., Bowers, J. in Munafo, R. M. (2014). Article Commentary: On the Persistence of Low Power in Psychological Science. *Quarterly Journal of Experimental Psychology*, 67, 1037–1040. <https://doi.org/10.1080/17470218.2014.885986>
64. Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C. in Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Annals of Internal Medicine*, 147(8), 573–577. <https://doi.org/10.7326/0003-4819-147-8-200710160-00010>
65. Xie, J. S. in Ali, M. J. (2023). To Slice or Perish. *Seminars in Ophthalmology*, 38(2), 105–107. <https://doi.org/10.1080/08820538.2023.2172813>
66. Yap, B. W. in Sim, C. H. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12), 2141–2155. <http://dx.doi.org/10.1080/00949655.2010.520163>

dr. Žiga Kozinc, doc.
Univerza na Primorskem
Fakulteta za vede o zdravju
ziga.kozinc@fvz.upr.si