

Občutljivost in specifičnost diagnostičnega testa

Občutljivost in specifičnost

Veliko populacijo $V = a + b + c + d$ pacientov, med katerimi je določen del okužen, testiramo. Testi imajo nekatere pomanjkljivosti. Tako test pri a okuženih potrди okužbo. Pri c okuženih pa test da negativen rezultat – to so *lažno negativni ali napačno negativni*. To imamo ponazorjeno v spodnji tabeli.

— — —	okužen	ni okužen
pozitiven test	a	b
negativen test	c	d

Povzemimo: a je število okuženih s pozitivnim testom, c je število okuženih, a z negativnim testom, torej lažno negativnih. In $a + c$ je število vseh okuženih.

Količnik

$$o = \frac{a}{a + c}$$

je **občutljivost** ali *senzitivnost* testa [4]. To je torej delež pozitivnih (na testu) med resnično okuženimi.

Pri neokuženih je test pri b osebah pozitiven – to so *lažno pozitivni* neokuženi. Pri d neokuženih pa je test pravilno dal negativen rezultat. Količnik

$$s = \frac{d}{b + d}$$

je **specifičnost** testa. To je torej delež negativnih (na testu) med tistimi, ki resnično niso okuženi.

Pomembno je, da vemo oba podatka; en sam je lahko močno zavajajoč. Imejmo populacijo, ki vsebuje tako okužene kot neokužene. Goljufigi test T, ki da vsakič pozitiven rezultat, ima stodontno občutljivost, saj je $c = d = 0$. Ima pa tudi ničelno specifičnost.

Ničvreden test U, ki da vsakič negativen rezultat, ima $a = b = 0$. Ni lažno pozitivnih, zato je specifičnost 100 % – občutljivost pa ničelna.

PCR test za COVID-19 zahteva več ur obdelave v laboratoriju, kjer pomnožujejo genetski material virusa in tako lahko zaznajo virus tudi pri majhnih koncentracijah. Po Wikipediji [8] (vsi podatki so iz januarja 2021) ima povprečno občutljivost okrog 95 %. Če smo bolj podrobni, naj bi bila

občutljivost nekako 70- do 100-odstotna, odvisno od znamke testa in metode odvzema vzorca.

Povprečna specifičnost naj bi bila okrog 99 % (od 92 do 100 odstotkov). Delež lažno pozitivnih med neokuženimi je zelo majhen. V skromnem deležu resnično okuženih pa test neupravičeno kaže odsotnost bolezni. PCR test velja za »zlati standard«: to je trenutno najbolj cenjen test na tržišču. Je eden od tako imenovanih *molekularnih* testov; razvijajo pa tudi nove, drugačne teste te kategorije.

Primer 1. Denimo, da imamo populacijo 10.000 ljudi, med katerimi je dva odstotka, torej 200 ljudi resnično okuženih. Pri občutljivosti 95 % test da pri $a = 190$ okuženih pozitiven rezultat. Deset oseb ($c=10$), torej en promil celotne populacije, je lažno negativnih. Nekateri med temi bodo brezskrbno hodili v službo ali družbo in morda okužili druge. Med 9800 neokuženimi in pri specifičnosti 99 % jih $d = 9702$ testira negativno, $b = 98$ pa lažno pozitivno. Ti zadnji bodo imeli deset dni izolacije in skrbi, čeprav niso okuženi. Nekateri med lažno pozitivnimi bodo morda verjeli, da so brez simptomov preboleli okužbo in da so torej imuni, čeprav to ni res.

V opisanem primeru je $b = 98$ oseb od $a + b = 288$ vseh oseb s pozitivnim testom v resnici brez okužbe. Torej je približno vsak tretji, ki ga je test označil kot okuženega, brez okužbe. Približno dve tretjini tistih s pozitivnim rezultatom pa je res okuženih.

Test je izločil izredno velik delež okuženih in če bodo ti res v karanteni, bo širjenje bolezni bistveno manjše. Velika večina neokuženih bo vsaj začasno pomirjena.

Primer 2. Podatki naj bodo kot zgoraj, le da je 30 odstotkov populacije, torej 3000 oseb okuženih. Pri 95-odstotni občutljivosti nam test pri $c = 150$ ljudeh da lažno negativen rezultat, $a = 2850$ oseb pa dobi pravilno diagnozo. Med 7000 neokuženimi je pri specifičnosti 0,99 en odstotek, torej $b = 70$ lažno pozitivnih, $d = 6930$ oseb pa dobi pravilno negativno diagnozo. Med vsemi ($a + b = 2920$ osebami) s pozitivnim rezultatom je le 70, torej približno 2,4 odstotka, lažno pozitivnih. Torej je med tistimi s pozitivnim testom približno 97,6 % res okuženih. Problem lažno negativnih pa je seveda bistveno večji kot v prejšnjem primeru.

Hitri antigenski testi

Hitri antigenski testi reagirajo na določeno, za virus značilno beljakovino na njegovi površini. Zato mora biti za pozitiven rezultat v vzorcu dovolj virusov. Časovni interval v poteku bolezni, v katerem test zazna virus, je

precej ožji kot pri PCR testu. Tako je občutljivost manjša kot pri PCR testih in je več lažno negativnih. Specifičnost pa naj bi bila enako visoka kot pri PCR testih. Tako lahko hitro identificiramo zelo kužne osebe. Ti testi so veliko cenejši in dajo rezultate že v četrt ali pol ure.

Svetovna zdravstvena organizacija (WHO) priporoča, naj uporabljamo hitre antigenske teste, ki imajo občutljivost vsaj 80 % in specifičnost vsaj 97 %. Evropska ustanova ECDC priporočilo za občutljivost zvišuje na vsaj 90 %. V Franciji gredo korak dalje in priporočajo specifičnost vsaj 99 %.

Problem je, če se pri izbiri testa zanašamo na podatke proizvajalcev. Kot pravi nedavni članek [2] v reviji *The Lancet*, je za izredno začasno odobritev testa v ZDA dovolj že, da producent predloži rezultate testiranja na 30 vzorcih. Preizkus je praktično zmeraj narejen na simptomatskih posameznikih (torej ljudeh s simptomi covid-19), ki so bili pozitivni na PCR testu (torej na osebah z visoko obremenitvijo z virusom). Kateri PCR test je bil to, proizvajalci ne navajajo. Nekateri proizvajalci v navodilih pravijo, da so hitri testi mišljeni le za simptomatske posameznike. Ti testi so novost na tržišču in zato ni veliko neodvisnih preverjanj občutljivosti in specifičnosti. Poleg tega so pri danem testu rezultati precej odvisni od izvedbe.

Raziskovalci z Univerze Oxford so v študiji [11] za hitri antigenski test I kalifornijskega podjetja ugotovili zelo visoko specifičnost: 99,6 %. Med vzorci, ki so bili na PCR testu pozitivni, pa je ta antigenski test v povprečju le 77 % spoznal za pozitivne. Če so teste opravljali laboranti, je bil ta delež nekaj višji (79 %), če priučeno medicinsko osebje, pa 73 %. Če so teste opravljali laiki (kot je denimo pri uporabi testov doma), je vrednost padla pod 58 %.

Kasnejša analiza [9] istega testa I, opravljena med množičnim testiranjem v Liverpoolu, je dala občutljivost 40 % in specifičnost 99,9 %. Med asimptomatskimi osebami z veliko produkcijo virusa je bila občutljivost 66 %.

Ameriška raziskava [5] hitrega antigenskega testa S na študentski populaciji (ki ni ravno odraz celotne populacije) je pokazala naslednje rezultate: na simptomatskih posameznikih: občutljivost 80 %, specifičnost 99 %; na asimptomatskih študentkah in študentih (brez znakov bolezni): občutljivost 41 %, specifičnost 98,4 %.

V francoski študiji [3] za ta isti test S na asimptomatskih osebah navajajo občutljivost 70 % in specifičnost 99 %. V dovoljenju za začasno uporabo testa S so navedeni rezultati iz proizvajalčeve študije: občutljivost 96,7 %, specifičnost 100 %.

Kako se znajti v džungli tako različnih, večkrat nasprotujočih si rezultatov za en in isti test? Kako izbrati dober test? Odgovor so *primerjalne raziskave*, kjer neodvisni znanstveniki več testov uporabijo pod istimi pogoji.

Nemški virologi in infektologi so v preprintu [1] preizkušali sedem anti-genskih testov, večinoma izdelkov velikih znanih podjetij. Pet jih je imelo specifičnost med 98,5 % in 100 %, eden je imel specifičnost 95 % in eden specifičnost 88 %. Test z najnižjo specifičnostjo je bil dobesedno premalo specifičen: zdi se, da je včasih reagiral tudi na viruse gripe in druge podobne viruse. Vendar je bilo s tem mogoče razložiti le pol lažno pozitivnih rezultatov. Namesto občutljivosti so raje merili, kakšne koncentracije virusa lahko test še zazna.

Francoska primerjava [3] desetih (verjetno tam najbolj popularnih) hitrih antigenih testov je pokazala na asimptomatskih posameznikih občutljivost od 30 do 50 odstotkov pri devetih testih. En sam test (S) je tu dosegel vrednost 70 %. Ta test je imel specifičnost 99 %, preostali pa 100 %.

Seznam hitrih antigenih testov, katerih rezultate so nedavno (15. januarja) priznavali pri vstopu v Francijo, je daljši in je na [10]. Med skoraj petdesetimi odobrenimi izdelki je tudi test podjetja Shenzhen Ultra-Diagnostics Biotec. Ta test je kupila naša država. Ni pa bil vključen v gornji primerjalni raziskavi. Kot razumemo, je pri nas večinoma uporabljen na asimptomatskih osebah.

Primer 3. V populaciji $V = 10.000$ asimptomatskih oseb je $x = 10$ %, torej 1000 oseb okuženih. Hitri antigeni test naj ima za to populacijo občutljivost 0,5 in specifičnost 0,99. Med 1000 okuženimi jih test odkrije 500. Preostalih 500 okuženih je lažno negativnih. Med 9000 neokuženimi je en odstotek, torej 90 lažno pozitivnih. Pravilno negativno diagnozo pa dobi 8910 oseb.

Test je torej izločil 500 okuženih in 90 neokuženih posameznikov. Prepustil pa je 500 okuženih.

Hitri antigeni testi spregledajo precejšen delež okuženih – predvsem tiste, pri katerih se bolezen šele razvija. Izločijo pa visok delež (trenutno) zelo kužnih posameznikov. Če tak test ponovimo po nekaj dneh, bomo izločili veliko tistih, pri katerih se je pri prvem testu bolezen šele začela.

V Beli hiši v ZDA so se zanašali na pogoste hitre antigenske teste, a je vseeno prišlo do izbruha bolezni. Treba je pač upoštevati tudi druge ukrepe za preprečevanje okužb.

Presejalni testi

Problem velikega deleža lažno pozitivnih imamo pri *presejalnih testih* za nekatere redke, a resne bolezni. To niso diagnostični testi. Gre za sito, uporabljeno na veliki populaciji, ki naj bi prepustilo kar se da velik delež oseb brez problemov in zadržalo praktično vse s problemom. Tako je preostanek

na situ (to so pozitivni na testu) bistveno manjši od testirane populacije. Na preostanku laže izvedemo dražja, zamudnejša presejanja ali bolj agresivne diagnostične postopke. Udeležencem takih testov je nujno treba pojasniti, da pozitiven presejalni test še ne pomeni diagnoze. (Tudi testiranje celotne populacije s hitrimi testi spada med presejalne teste.)

Eden od presejalnih testov bazira na rentgenskih slikah. Razlike med kakovostjo dela radiologov so lahko precejšnje. Branje komaj vidnih lis na sliki ni enostavno in zahteva veliko pozornosti in izkušenj. Tudi v literaturi se podatki o senzitivnosti in specifičnosti danega testa precej razlikujejo.

Če se trudimo, da ne bi spregledali tudi najmanjše potencialno problematične spremembe, se pravi, povečujemo občutljivost, tvegamo več lažno pozitivnih rezultatov, torej zmanjšanje specifičnosti. Uravnoteženje občutljivosti in specifičnosti je pogosto delikatna naloga.

Primer 4. Privzemimo, da občutljivost in specifičnost presejalnega testa znašata 90 odstotkov, torej $s = o = 0,9$. Denimo, najprej, da je v populaciji delež nediagnosticiranih *problemov* (kar ustreza *okuženim* v prejšnjih primerih) en odstotek. Pri $V = a + b + c + d = 10.000$ testiranih ima le $a + c = 100$ oseb problem. Med njimi bo $a = 90$ dobilo pravo diagnozo, $c = 10$ pa lažno negativen rezultat. Od 9900 neproblematičnih pacientk (pacientov) bi $b = 990$ dobilo lažen pozitiven rezultat, 8910 pa pravilen negativen rezultat. Na situ je od 10.000 oseb ostalo 1080 ljudi. Med njimi jih ima le 90 problem, torej le ena dvanaajstina ali približno 8,3 odstotka. Lažno pozitivnih je enajstkrat toliko, kot je resnično ogroženih. Skozi sito je padlo 10 oseb s problemom in ti za zdaj ne bodo deležni nadaljnje obravnave.

Pozitivna in negativna napovedna vrednost

Iz primerov vidimo, da za uspešnost oziroma uporabnost testa nista pomembni le njegova občutljivost in specifičnost, temveč tudi delež oseb v populaciji, ki imajo problem. Naj bo $V = a + b + c + d$ celotna populacija. Označimo z

$$x = \frac{a + c}{V}$$

delež okuženih (problematičnih) oseb v testu.

Torej je $a + c = xV$ in $b + d = V - xV = (1 - x)V$. Vemo, da je $a = o(a + c) = oxV$ in $d = s(b + d) = s(1 - x)V$. Od tod je $b = b + d - d = (1 - x)V - s(1 - x)V = (x(s - 1) + 1 - s)V$ in $c = a + c - a = xV - oxV = (x - ox)V$.

Vpeljali bomo dve novi merili. To sta *pozitivna napovedna vrednost* in *negativna napovedna vrednost*. Obe ležita na intervalu od 0 do 1 in sta močno odvisni od deleža x okuženih (problematičnih) oseb.

Pozitivna napovedna vrednost ali *preciznost* p testa je za našo tabelo enaka

$$p = \frac{a}{a + b}.$$

Število okuženih s pozitivnim testom delimo s številom vseh, ki imajo pozitiven test. Po gornjih računih je

$$p(x) = \frac{ox}{(s + o - 1)x + 1 - s}.$$

Tako je p ulomljena linearna funkcija spremenljivke x . Pri fiksni občutljivosti o in specifičnosti s je p močno odvisen od deleža okuženih, saj je $p(0) = 0$, $p(1) = 1$.

Pri presejalnih testih je delež x problematičnih oseb v populaciji blizu 0. Ker je $p(0) = 0$, je pozitivna napovedna vrednost $p(x)$ majhna: na situ ostane preveč oseb. To je velik problem tako za paciente (pacientke) kot za javno zdravstvo. Če so za napačne rezultate odgovorne slučajne napake (šum), pomaga ponovitev presejalnega testa. Recimo, posamezna dodatna digitalna rentgenska slika je glede izpostavljenosti sevanju malenkost v primerjavi z računalniško tomografijo (CT), ki je ekvivalent nekaj sto posameznim slikam. Mogoče so dodatne nenevarne preiskave, kot je ultrazvok. Včasih pomaga, če originalno sliko neodvisno pregledata dva strokovnjaka, ki ob nestrinjanju pokličeta še tretjega. Pomembno je, da zmanjšamo število nadaljnjih bolj agresivnih diagnostičnih postopkov.

Negativna napovedna vrednost n testa za našo tabelo je število neokuženih z negativnim testom deljeno s številom vseh, ki imajo negativen test:

$$n = \frac{d}{c + d}.$$

Ker je $c = (x - ox)V$ in $d = s(1 - x)V$, je

$$n(x) = \frac{s(1 - x)}{s - (s + o - 1)x}.$$

Vidimo: $n(0) = 1$, $n(1) = 0$. Če je delež x problematičnih (okuženih) majhen, pričakujemo, da bo $n(x)$ blizu 1 in torej negativna napovedna vrednost visoka.

Oglejmo si še enkrat primer 1. Pozitivno napovedno vrednost smo že izračunali: $p = 190/288 \approx 66\%$. To odraža dejstvo, ki smo ga že povedali: med pozitivno testiranimi je približno dve tretjini res okuženih.

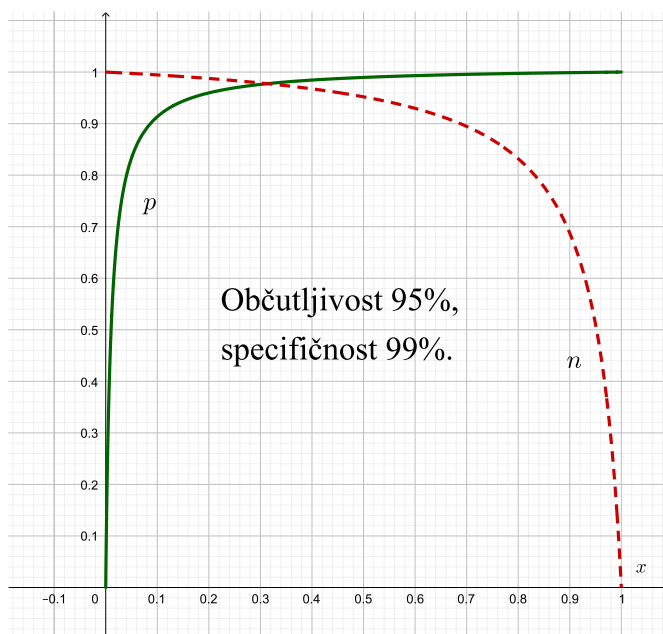
Med 9712 osebami z negativnim testom je 9702 res neokuženih, torej je negativna napovedna vrednost izredno visoka: $n \approx 99,9\%$. Za posameznika je pri negativnem testu možnost okužbe skoraj izključena.

Občutljivost in specifičnost diagnostičnega testa

V primeru 2 je pri isti občutljivosti in isti specifičnosti kot v prvem primeru: $p = 2850/(2850 + 70) \approx 97,6 \%$. Tolikšen delež je res okuženih med vsemi, ki so bili na testu pozitivni. Pozitivna napovedna vrednost je zdaj visoka. Pozitiven rezultat na testu skoraj gotovo pomeni okužbo.

Med 7080 osebami z negativnim testom je 6930 res neokuženih, torej je negativna napovedna vrednost prav tako visoka: $n \approx 98 \%$.

Na sliki 1 imamo grafa za pozitivno in negativno napovedno vrednost (črtkano) pri $o = 0,95$, $s = 0,99$.



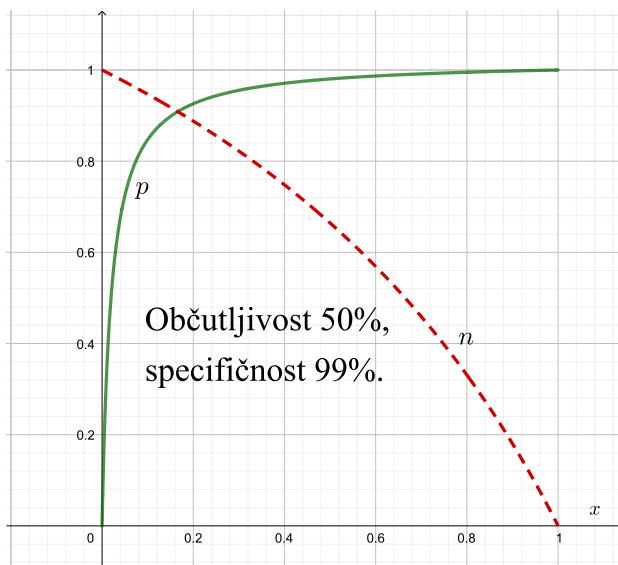
Slika 1. Pozitivna in negativna napovedna vrednost za $o = 0,95$, $s = 0,99$.

V primeru 3 (antigenski test) je med 590 osebami s pozitivnim testom 500 res okuženih, torej je pozitivna napovedna vrednost $500/590 \approx 85 \%$.

Med 9410 ljudmi z negativnim testom je 8910 res neokuženih, zato je negativna napovedna vrednost $8910/9410 \approx 95 \%$.

V našem primeru je bilo med 10.000 testi 590 pozitivnih, torej slabih 6 odstotkov. Delež okuženih v testirani populaciji pa je višji in znaša 10 odstotkov. Tudi to ilustrira pomanjkljivosti testa.

Na sliki 2 sta grafa za pozitivno in negativno napovedno vrednost (črtkano) pri $o = 0,5$, $s = 0,99$.



Slika 2. Pozitivna in negativna napovedna vrednost za $o = 0,5$, $s = 0,99$.

V četrtem primeru (presejalni test) smo že izračunali, da je pozitivna napovedna vrednost zelo nizka in enaka $1/12 \approx 8,3$ %. Kljub temu bi lahko rekli, da je test opravil svojo nalogo: populacija, ki gre v nadaljnjo obravnavo, je približno ena devetina prvotne.

Negativna napovedna vrednost pa je izredno visoka in enaka $8910/8920 \approx 99,9$ %. To pač pomeni, da si pri negativnem testu lahko oddahnemo: verjetnost, da imamo problem, je izredno majhna.

Če v četrtem primeru povečamo vrednost za x na 5 odstotkov, je pozitivna napovedna vrednost enaka $450/(450 + 950) \approx 32$ %. V tem primeru ima približno vsaka tretja oseba s pozitivnim testom res problem. Pozitivna napovedna vrednost je zdaj mnogo boljša. Negativna napovedna vrednost je enaka $8550/(8550 + 50) \approx 99,4$ % in tako še vedno izredno visoka.

Naloga. V raziskavi [7] je bilo testiranih 122 oseb. Izračunali so, da je bila občutljivost testa 97 %, specifičnost 64,5 %, pozitivna napovedna vrednost 89 %. Približno koliko oseb je v resnici imelo problem?

Pisanje tega prispevka je spodbudilo branje odlične poljudne knjige *The Maths of Life and Death, Why Maths Is (Almost) Everything* [6]. Napisal jo je Kit Yates, profesor in eden od direktorjev Centra za matematično biologijo

na Univerzi v angleškem mestu Bath. Čeprav je knjiga izšla pred sedanjo pandemijo, je njena vsebina še kako relevantna v današnjih razmerah in nam je bila za vzor, skupaj s predavanji [4] profesorice dr. Maje Petek Šter. Z veseljem smo uporabili pripombe obeh recenzentov.

(Rešitev naloge: 91–92.)

LITERATURA

- [1] V. C. Corman et al., *Comparison of seven commercial SARS-CoV-2 rapid Point-of-Care Antigen 1 tests*, preprint, 2020, dostopno na www.medrxiv.org/content/10.1101/2020.11.12.20230292v1.full.pdf, ogled 25. 1. 2021.
- [2] M. C. Fitzpatrick et al., *Buyer beware: inflated claims of sensitivity for rapid COVID-19 tests*, *The Lancet* 397(10268) (2021) 24–25, dostopno na www.thelancet.com/action/showPdf?pii=S0140-6736, ogled 25. 1. 2021.
- [3] S. Fourati et al., *Évaluation de la performance diagnostique de neuf tests rapides antigéniques COVID-19*, Assistance publique – Hôpitaux de Paris, 2020, dostopno na www.aphp.fr/contenu/evaluation-de-la-performance-diagnostique-de-neuf-tests-rapides-antigeniques-covid-19, ogled 25. 1. 2021.
- [4] M. Petek Šter, *Interpretacija rezultatov statističnih testov*, Praktična statistika, Medicinska fakulteta, Ljubljana, februar–marec 2016, dostopno na www.mf.uni-lj.si/application/files/9315/3842/0966/EBM_petek_2.pdf, ogled 25. 1. 2021.
- [5] I. W. Pray et al., *Performance of an Antigen-Based Test for Asymptomatic and Symptomatic SARS-CoV-2 Testing at Two University Campuses – Wisconsin*, 2020, CDC Morbidity and mortality weekly report, 1. jan 2021 / 69(5152); 1642–1647, dostopno na www.cdc.gov/mmwr/volumes/69/wr/mm695152a3.htm, ogled 25. 1. 2021.
- [6] K. Yates, *The Maths of Life and Death, Why Maths Is (Almost) Everything*, Quercus, London, 2019.
- [7] M. Zeeshan et al., *Diagnostic Accuracy of Digital Mammography in the Detection of Breast Cancer*, *Cureus* 10(4), 2018, dostopno na www.researchgate.net/publication/324312923_Diagnostic_Accuracy_of_Digital_Mammography_in_the_Detection_of_Breast_Cancer, ogled 25. 1. 2021.
- [8] *COVID-19 testing*, Wikipedia, dostopno na en.wikipedia.org/wiki/COVID-19_testing, ogled 25. 1. 2021.
- [9] *Expert reaction to Interim Evaluation Report from the Liverpool Covid-19 Community Testing Pilot*, Science Media Centre, 2020, dostopno na www.sciencemediacentre.org/expert-reaction-to-interim-evaluation-report-from-the-liverpool-covid-19-community-testing-pilot/, ogled 25. 1. 2021.
- [10] *List of antigen tests authorized for entry to France from UK*, dostopno na uk.ambafrance.org/List-of-antigen-tests-authorized-for-entry-to-France-from-UK-29434, ogled 25. 1. 2021.
- [11] *Preliminary report from the Joint PHE Porton Down and University of Oxford SARS-CoV-2 test development and validation cell: Rapid evaluation of Lateral Flow Viral Antigen detection devices (LFDs) for mass community testing*, dostopno na www.ox.ac.uk/sites/files/oxford/media_wysiwyg/UK%20evaluation_PHE%20Porton%20Down%20%20University%20of%20Oxford_final.pdf, ogled 25. 1. 2021.

Peter Legiša