

MNOŽIČNI PODATKI V FARMACIJI

BIG DATA IN PHARMACY

AVTORJA / AUTHORS:

asist. Janja Jazbar, mag. farm.
prof. dr. Mitja Kos, mag. farm.

*Univerza v Ljubljani, Fakulteta za farmacijo,
Katedra za socialno farmacijo,
Aškerčeva 7, 1000 Ljubljana*

NASLOV ZA DOPISOVANJE / CORRESPONDENCE:

E-mail: janja.jazbar@ffa.uni-lj.si

1 UVOD

V zadnjih 24 urah je na svetu nastalo 2,5 eksabajtov podatkov (eksa pomeni 10^{18}) (1). Tako velike količine podatkov si težko predstavljamo, v več virih pa lahko preberemo, da je 90 % vseh podatkov, ki so na voljo, nastalo v zadnjih dveh letih. Povedano še drugače, v vseh tiskanih knjigah na svetu je samo 6 % vseh podatkov, 94 % pa jih shranjujemo v digitalni obliki (1). Ogromne količine podatkov, ki se generirajo vsak dan, odpirajo nepredstavljive priložnosti na različnih področjih, vse od ciljanih oglasov in napovedovanja vremena pa do odkrivanja pranja denarja in raziskav globalnega segrevanja. Tudi na področju farmacije se vsakodnevno ustvarjajo podatki, ki lahko z ustrezno analizo prinašajo pomembno korist za pacienta in zdravstveni sistem. Pri tem pa se bo treba soočiti s številnimi izzivi, ki jih t. i. množični podatki (*big data*) prinašajo.

POVZETEK

V vseh tiskanih knjigah na svetu je samo 6 % vseh podatkov, 94 % pa jih shranjujemo v digitalni obliki. Izraz *množični podatki* (*big data*) opredeljuje »zelo velike zbirke podatkov, ki so lahko kompleksne, multi-dimenzionalne, nestrukturirane in heterogene, ki se zelo hitro akumulirajo in ki jih lahko računalniško analiziramo, da odkrijemo vzorce, trende in povezave«. Za analizo množičnih podatkov potrebujemo napredne in specializirane metode analize podatkov. Tudi na področju farmacije se vsakodnevno ustvarjajo podatki, ki lahko z ustrezno analizo prinašajo pomembno korist za pacienta in zdravstveni sistem: hitrejši razvoj zdravil, bolj učinkovita in varna terapija, prilagojena posamezniku, boljša farmakovigilanca, bolj učinkovit in varčen zdravstveni sistem. Vendar kot vse najbolj napredne ideje v znanosti, tudi množični podatki ustvarjajo številne etične dileme. To velja še posebej za področji medicine in farmacije, kjer podatki lahko zajemajo občutljive informacije. Analiza množičnih podatkov lahko revolucionarno spremeni način zdravljenja, pri vpeljavi novosti s tega področja v zdravstveni sistem in prakso pa bo potrebna posebna pozornost pri različnih izzivih, ki jih taka metodologija prinaša.

KLJUČNE BESEDE:

farmacija, medicina, farmakoepidemiologija, množični podatki

ABSTRACT

All books ever printed in the world contain only 6% of all data, whereas 94% of data are stored in digital format. Big data are defined as »extremely large datasets which may be complex, multi-dimensional, unstructured and heterogeneous, which are accumulating rapidly and which may be analysed computationally to reveal patterns, trends, and associations«. Analysis of big data require advanced or specialised methods. New data are generated every day also in the field of pharmacy. Appropriate use and analysis of big data could bring important benefits to individual patients and health system in general: faster medication development, more effective and safe pharmacotherapy with personalised medicine, better pharmacovigilance, more effective and cost-effective health system. However, several



ethical aspects have been considered regarding the use of big data, especially in the field of medicine and pharmacy, as the data can contain sensitive personal information. Big data analytics has the potential to revolutionise medicine, nevertheless several challenges should be considered when implementing new technologies in practice.

KEY WORDS:

Big data, medicine, pharmacoepidemiology, pharmacy

2 OPREDELITEV POJMA MNOŽIČNI PODATKI

Na področju informatike imata izraza podatek in informacija različen pomen. Podatek je opredmeteno dejstvo, torej črke, številke ali znaki, s katerimi predstavimo neko dejstvo. Informacijo pa dobimo, ko podatku dodamo pomen. Informacija torej nastane šele z interpretacijo podatkov in se seveda lahko razlikuje glede na predznanje posameznika, ki podatek interpretira. Množične podatke tako sestavljajo ogromne količine podatkov, bistvena pa je njihova pravilna interpretacija, ki nam poda prave in uporabne informacije. Področje množičnih podatkov se šele razvija in v skladu s tem se spreminja tudi definicija. Definicija iz leta 2019 opredeli množične podatke kot »zelo velike zbirke podatkov, ki so lahko kompleksne, multi-dimenzionalne, nestrukturirane in heterogene, ki se zelo hitro akumulirajo in ki jih lahko računalniško analiziramo, da odkrijemo vzorce, trende in povezave. Običajno za analizo množičnih podatkov potrebujemo napredne in specializirane metode analize podatkov« (2). Pogosto namesto klasične definicije opredelimo značilnosti množičnih podatkov: količina (*volume*), hitrost (*velocity*), raznolikost (*variety*), verodostojnost (*veracity*) in vrednost (*value*) (3–5). Temu različni avtorji dodajajo še druge značilnosti, npr. variabilnost (*variability*), vizualizacija (*visualisation*), varnost (*vulnerability*) (6) itd. Naštete značilnosti množičnih podatkov v povezavi s področjem farmacije so predstavljene na sliki 1 in razložene v nadaljevanju.

Količina (*volume*): Izraz množični podatki pomeni tako velike količine podatkov, da jih s klasičnimi orodji ne moremo obdelovati. Čeprav stroga meja ni postavljena, gre pri tem za svet terabajtov, petabajtov, eksabajtov in zetabajtov (7). En eksabajt predstavlja milijardo gigabajtov. Celotna kumulativna svetovna količina zdravstvenih podatkov naj bi

s 150 eksabajtov v letu 2013 narasla na preko 2300 eksabajtov v letu 2020 (8). Vsak trenutek po svetu poteka okoli 100.000 kliničnih raziskav in vsako leto v Pubmedu objavijo 1,8 milijona novih člankov. Posledično je nemogoče slediti vsemu znanju že samo na področju biomedicine. Po nekaterih analizah traja približno 17 let, da določena inovacija doseže klinično prakso – ta časovni okvir bi se lahko z ustrezno uporabo množičnih podatkov občutno zmanjšal (8). Tako velike količine podatkov zahtevajo posebno tehnologijo za shranjevanje in prenos podatkov, npr. shranjevanje v oblaku, in kompleksne metode analize podatkov, kot so podatkovno rudarjenje, strojno učenje in umetna inteligenca (6).

Hitrost (*velocity*): Poleg količine podatkov dodaten izziv predstavlja tudi hitrost zbiranja podatkov, ki so danes lahko generirani tudi v realnem času (*real-time*). Velikim administrativnim zbirkam podatkov, ki se zbirajo na nacionalni ravni, se pridružujejo še podatki rednega merjenja zdravstvenega statusa na klinikah ali preko mobilnih aplikacij, npr. dnevno merjenje ravni glukoze pri bolnikih s sladkorno boleznijo, redne meritve krvnega tlaka ali EKG. Istočasno se povečuje tudi hitrost, ki je zahtevana za analizo podatkov in primerjavo podatkov ter sklepanje odločitev na podlagi rezultatov takih analiz podatkov (4). V prihodnosti bi lahko s pomočjo podatkov v realnem času veliko hitreje zaznali širjenje nalezljivih bolezni, jih hitreje identificirali in odpravili. Na ta način bi lahko pomembno zmanjšali obolenost in preprečili izbruhe v bolnišnicah (4).

Raznolikost (*variety*): Večina zdravstvenih podatkov, ki so danes na voljo za obdelavo, so strukturirani podatki. Strukturirani podatki so tisti, ki so oblikovani in urejeni v polja in preglednice. Strukturirane podatke iz nacionalnih zbirk je lažje shranjevati in analizirati. Takim podatkom se v zadnjih letih priključujejo tudi semistrukturirani in nestrukturirani podatki. Primer slednjih so slike, videi ali nestrukturiran tekst iz mobilnih aplikacij ali socialnih medijev. Razvoj k posameznemu bolniku usmerjene medicinske obravnave (*precision medicine*) zahteva dostop in integracijo zelo različnih vrst podatkov, od tradicionalnih kliničnih podatkov, do laboratorijskih vrednosti, genetskih podatkov, posnetkov (*imaging*) in podatkov, ki jih sporočajo pacienti (8). Zaenkrat shranjujemo in obdelujemo samo majhen del takih podatkov. Potencial množičnih podatkov v zdravstvu se skriva v pravi kombinaciji strukturiranih, semistrukturiranih in nestrukturiranih podatkov. Cilj je torej uporaba tako tradicionalnih oblik zbiranja podatkov kot tudi novih multimedialskih podatkov v končno korist pacienta (4).

Verodostojnost (*veracity*): Velika količina podatkov seveda ne zagotavlja nujno verodostojnih informacij (9). Eden od

takih primerov je precej napoved prevalence gripe v letu 2013 v ZDA na podlagi iskalnih profilov uporabnikov interneta (10). Za veljavnost rezultatov pri analizi množičnih podatkov je običajno nujen pogoj kakovost vhodnih podatkov. Pogosto podatke, ki ustvarjajo množične podatke, zbiramo za različne namene, ponavadi ne za namen analize določenega raziskovalnega vprašanja. Pri analizi velike količine podatkov obstaja tudi možnost lažnih alarmov in lažnih povezav, ko ugotovljena statistična povezava ne pomeni nujno, da sta določen dejavnik in izid vzročno povezana. Verodostojnost množičnih podatkov je torej cilj, ni pa vedno realnost. Kakovost podatkov je v zdravstvu še toliko bolj pomembna, saj lahko rezultat analize podatkov vpliva na odločitev o strategiji zdravljenja, kar lahko bistveno vpliva na izid zdravljenja, med drugim tudi npr. prepreči smrt. Izboljšanje zdravstvene oskrbe, zmanjšanje zdravstvenih na-

pak in napredek pri učinkovitosti in varnosti zdravil so odvisni od kakovosti in pravilnosti vhodnih podatkov. S tem je povezano tudi zaupanje v rezultate analiz podatkov. Prav na področju zdravstva bo verodostojnost informacij v prihodnosti ključna za vpeljavo novih sistemov v prakso (4, 10).

Vrednost (*value*): Velika količina podatkov sama po sebi nima nobene vrednosti, če ne znamo iz podatkov razbrati uporabnih informacij. Izraz vrednost se nanaša na to, kako rezultate analiz podatkov uporabimo pri odločanju v praksi. Posamezni viri podatkov imajo lahko manjšo vrednost, ko pa podatke zbiramo in povezujemo iz različnih virov in jih obdelamo z ustrežno analizo ter testiramo tudi verodostojnost podatkov, potem lahko rezultati dobijo pravo vrednost in prinašajo vpogled v določeno področje in pomembna nova spoznanja (3).



Slika 1. Najpogostejše opredeljene značilnosti množičnih podatkov – 5 V-jev. Najprej so množične podatke opredelili s tremi značilnostmi – količina, hitrost in raznolikost. Kasneje so temu dodali še dva pojma – verodostojnost in vrednost. V zadnjih letih se temu pridružujejo še številne druge značilnosti. Vir ikon na sliki: Flaticon.com.

Figure 1. Most commonly defined characteristics of Big Data – 5 V's of Big Data. At first, Big data were described with three characteristics – volume, velocity and variety. Two additional characteristics were later included – veracity and value. Several other characteristics have been added lately. Icons by Flaticon.com.

3 PRILOŽNOSTI ZA UPORABO MNOŽIČNIH PODATKOV NA PODROČJU FARMACIJE

Učinkovita analiza množičnih podatkov lahko na področju farmacije veliko doprinese, tako pacientom kot tudi zdravstvenim delavcem in ustanovam, od lekarn do zdravstvenih domov in bolnišnic in na koncu tudi celotnemu zdravstvenemu sistemu (4). Množične podatke na področju zdravstva v grobem razdelimo v tri skupine. Najprej so to medicinski podatki o zdravstvenem stanju posameznika (npr. podatki iz elektronske kartoteke, laboratorijske vrednosti, genetski podatki in radiološki posnetki), temu sledijo zdravstveni podatki iz velikih zbirk podatkov (npr. podatki o izdanih receptih, hospitalizacije, podatki o onesnaženosti zraka

in znanstvena literatura), nenazadnje pa so na voljo tudi podatki, ki jih poročajo pacienti (npr. podatki mobilnih aplikacij ali socialnih medijev, ki jih sporočajo pacienti ali ožji družinski člani). Različni potencialni viri zdravstvenih podatkov so predstavljeni na sliki 2.

V znanstveni literaturi omenjajo številna področja, kjer naj bi analiza množičnih podatkov v prihodnosti prinesla pomemben napredek (4). Najprej omenimo področje javnega zdravja in epidemiologije, saj smo v preteklem letu lahko vsi spremljali uporabo množičnih podatkov v pandemiji covid-19. Za pravočasno in ustrezno ukrepanje ob širjenju nalezljivih bolezni je zelo pomembna hitra in ustrezna analiza podatkov. Projekt Covid-19 Sledilnik (11) – projekt velike skupine prostovoljcev – zbira, analizira in objavlja podatke o širjenju koronavirusa SARS-CoV-2 v Sloveniji v realnem času. Taki podatki so nepogrešljivi za pravočasne javno-



Slika 2: Viri zdravstvenih podatkov, ki lahko prispevajo k množičnim podatkom na področju farmacije. Vir ikon na sliki: Flaticon.com.
Figure 2: Different sources of health data that contribute to big data in the field of pharmacy. Icons by Flaticon.com.

zdravstvene ukrepe. Jure Leskovec in sodelavci z Univerze Stanford (12) so uporabili množične podatke o mobilnosti populacije v modelu za simulacijo širjenja epidemije koronavirusa SARS-CoV-2 v ZDA. Rezultati njihovih raziskav kažejo, kateri ukrepi so pri spopadanju z epidemijo lahko bolj in kateri manj uspešni (12). Prav preko epidemije covid-19 smo naredili veliko napredka pri uporabi množičnih podatkov v javnem zdravju, pokazale pa so se tudi pomanjkljivosti, ki jih bo potrebno v prihodnosti nasloviti.

Tudi na številnih drugih področjih medicine in farmacije se obetajo spremembe s pomočjo analize množičnih podatkov. Na področju primerjalne učinkovitosti lahko s pomočjo analize množičnih podatkov določimo bolj klinično relevantne in stroškovno učinkovite načine zdravljenja pacientov. Analiza množičnih podatkov omogoča hitrejši in bolj tarčno usmerjen razvoj zdravil in medicinskih pripomočkov ter bolj učinkovito načrtovanje kliničnih raziskav. Z dokazi podprta medicina bo lahko z analizo množičnih podatkov zelo napredovala in bo omogočala identifikacijo najbolj učinkovite in varne terapije za individualnega bolnika ter bolj učinkovito preventivo oz. usmeritev terapije in preventive k najbolj ranljivim posameznikom. Prihodnost so tudi genske analize kot del vsakdanje klinične prakse za optimalno zdravljenje individualnega bolnika. Množični podatki omogočajo hitrejšo analizo zavarovalniških prevar in detekcijo ponarejenih zdravil. Obetajoča je tudi analiza podatkov na področju farmakovigilance in varnosti zdravil, kjer bi lahko v realnem času zajemali podatke iz klinik in mobilnih ali e-aplikacij za večjo varnost pri uporabi zdravil in ugotavljanje neželenih dogodkov (4).

V nadaljevanju predstavljamo štiri izbrane primere možnosti uporabe množičnih podatkov na področju farmacije.

Farmakoepidemiologija: Farmakoepidemiologija je področje, ki proučuje učinkovitost in varnost zdravil po pridobitvi dovoljenja za promet (13). S pomočjo opazovalnih raziskav in uporabe zdravstvenih podatkov, ki se rutinsko zbirajo, farmakoepidemiologija proučuje učinke terapij v realnem okolju (npr. ob sočasni uporabi drugih zdravil), dolgoročne želene in neželene učinke terapij, zelo redke neželene učinke, učinke terapij v občutljivih populacijah, kot so starejši, otroci ali nosečnice, ali npr. varnost in učinkovitost zdravil pri neodobreni uporabi zdravil (*off-label use*) (13, 14). To je le nekaj pomembnih prednosti, ki jih prinašajo farmakoepidemiološke raziskave v primerjavi s kliničnimi raziskavami, pri interpretaciji rezultatov pa je potrebno upoštevati tudi omejitve opazovalnih raziskav, ki med drugim vključujejo potencialne moteče dejavnike in druge pristranosti, manjkajoče podatke in zahtevno določanje vzročnosti povezav. V zadnjem desetletju se je to

področje zelo razvilo, pričakujemo pa še hitrejši razvoj v prihodnosti z uporabo metod strojnega učenja. Primer aplikacije analize množičnih podatkov na področju farmakoepidemiologije v prihodnosti je napovedovanje odziva na terapijo pri bolnikih z revmatoidnim artritisom, ki je trenutno v klinični praksi velik izziv (13).

Neželeni dogodki: Neželeni dogodki so vzrok za pomemben del obolevnosti in celo pojavov smrti, velikokrat pa jih je možno preprečiti. Z ustrezno analizo podatkov bi lahko predvideli, kateri pacienti imajo največje tveganje za pojav neželenih dogodkov. Primer je ledvična odpoved zaradi kombinacije zdravil. Odpoved ledvic ima veliko tveganje za smrt, prinaša pa tudi veliko stroškov. Meritve ledvične funkcije lahko zaznajo zgodnje spremembe v delovanju ledvic, preden se pojavi ledvična odpoved. V prihodnosti bi lahko s pomočjo kompleksne analize podatkov v realnem času kombinirali podatke o meritvah ledvične funkcije, terapiji z zdravili, krvnem tlaku in ostalih pomembnih parametroh in na ta način identificirali posameznike s tveganjem za ledvično odpoved dovolj zgodaj, da bi lahko ta neželeni učinek preprečili. Na podoben način bi lahko s pomočjo genetskih in genomskih podatkov, laboratorijskih podatkov, vitalnih znakov in podatkov o zdravljenju predvideli tudi druge neželene dogodke zaradi zdravil pri individualnem pacientu (15).

K posameznemu bolniku usmerjena medicinska obravnava (*precision medicine*): Ta pojem se nanaša na prilagajanje terapije individualnim značilnostim vsakega pacienta. V idealnem primeru torej k posameznemu bolniku usmerjena medicinska obravnava omogoča odločitve o najboljši, torej najbolj učinkoviti in najbolj varni terapiji za vsakega pacienta, kar prinaša tudi prihranke v celotnem zdravstvenem sistemu. Zaenkrat je to področje zelo omejeno na posamezne značilnosti pacienta, kot je določena genetska informacija. Zanimivo dejstvo je npr. to, da je nedavni projekt *1000 Genoms* v šestih mesecih dela v bazo podatkov *NCBI GenBank* shranil dvakrat več podatkov, kot jih je bilo do takrat shranjenih v tej bazi v preteklih 30 letih (16). V zadnjem desetletju so objavili številne genomske markerje, ki določajo učinkovitost zdravil, odmerjanje zdravil ali možnost neželenih učinkov. Na področju farmacije so najbolj znani zagotovo primeri določanja genotipov genov posameznih encimov iz družine citokromov P450, ki omogočajo individualno prilagajanje odmerka zdravila; primer je genetski polimorfizem v genih CYP2C9 in VKORC1 in odmerjanje varfarina. Zaenkrat te podatki še zelo redko uporabljamo v klinični praksi. V prihodnosti bi lahko temu dodali tudi druge multimedijske podatke, kar bi omogočalo optimalno izbiro zdravil in odmerjanje zdravil (17).

Razvoj zdravil: Razvoj zdravil poteka v več korakih, od identifikacije tarče, presejanja molekul ter identifikacije in optimizacije spojine vodnice, do predkliničnih in kliničnih preskušanj, razvoja formulacije ter regulatornih postopkov za odobritev zdravila, čemur sledijo kasneje še raziskave po utrženju zdravila. Celoten proces prinaša veliko stroškov, kar se odraža v končni ceni zdravila. Racionalno načrtovanje učinkovin že vključuje npr. virtualno reševanje, kar skrajša postopek in zmanjša stroške. Kljub temu je na tem področju še veliko maneverskega prostora za učinkovito uporabo analize množičnih podatkov tako v začetnih fazah razvoja spojine vodnice kot tudi kasneje v načrtovanju kliničnih raziskav za bolj učinkovit, hitrejši in cenejši razvoj zdravil (18, 19).

4 ETIKA IN MNOŽIČNI PODATKI

Najbolj napredne ideje v znanosti pogosto ustvarijo številne etične dileme, ki jih je potrebno na novo opredeliti. To velja še posebej za področje farmacije, kjer se obdelujejo mnogi najbolj občutljivi podatki.

Verjetno je na področju biomedicine ena najbolj prepoznanih dilem povezana z dovoljenjem posameznika za uporabo njegovih podatkov v namen neke raziskave oz. t. i. informirani pristanek (20, 21). V tradicionalni raziskavi vsak udeleženec podpiše informirani pristanek, s katerim se strinja z uporabo njegovih podatkov za namen konkretne raziskave. Množični podatki so že v osnovi namenjeni raziskovanju nepredvidljivih povezav, zato klasičen informirani pristanek ne pride v poštev. Poleg tega so s tem povezane tudi dileme ponovne uporabe podatkov za različne namene in povezovanje podatkov med različnimi državami, kjer velja različna zakonodaja. Druga najbolj pogosta etična dilema pa je povezana z zasebnostjo podatkov, skupaj z anonimnostjo in zaupnostjo podatkov (20). Podatki, ki so v javnosti, torej npr. na internetu, niso hkrati nujno tudi javni podatki. Pogosto se zaupnost podatkov povezuje z anonimnostjo in posledično s potrebo po informiranem pristanku. To področje deloma ureja zakonodaja posamezne države, deloma pa skrbniki konkretnih javnih in zasebnih zbirk podatkov. Zaenkrat splošno sprejetih rešitev za zgornje etične izzive še ni (20, 21).

Lastništvo je kompleksna etična dilema, ki se tiče pravic do obdelave podatkov ter po drugi strani koristi od intelektualne lastnine in inovacij kot posledice analize množičnih

podatkov. Strokovnjaki in znanstveniki kot pomembno dilemo omenjajo tudi epistemologijo – izraz se nanaša na dejstvo, da je analiza množičnih podatkov tako kompleksna stvar, ki je večina laikov niti strokovnjakov s področja biomedicine ne more popolnoma razumeti, zato bo izredno težko validirati rezultate takih analiz podatkov. Etično sporno je lahko tudi dejstvo, da si vsaj zaenkrat samo največja in najbogatejša podjetja lahko privoščijo shranjevanje in analizo množičnih podatkov, kar omejuje tako raziskovanje kot tudi uporabo rezultatov samo za določeno populacijo. Še številne druge etične dileme se pojavljajo in jih bo potrebno nasloviti za uspešen prehod analize množičnih podatkov v prakso (20, 21).

5 SKLEP

Analiza množičnih podatkov lahko revolucionarno spremeni način zdravljenja, kot ga poznamo danes. Razvoj zdravil se lahko pospeši, zdravljenje in preventiva lahko postaneta veliko bolj prilagojena posamezniku in posledično bolj učinkovita in bolj varna, zdravstveni sistem pa lahko pridobi tako na učinkovitosti kot tudi na zmanjšanju stroškov. Pri vpeljavi novosti s področja analize množičnih podatkov v zdravstveni sistem in prakso pa bo potrebna posebna pozornost pri različnih izzivih, ki jih taka metodologija prinaša. Človeške zmogljivosti same že dolgo niso več kos napredku v farmaciji, po drugi strani pa tudi sami izsledki, pridobljeni na podlagi računalniške analize podatkov, niso dovolj in za pravilno uporabo potrebujejo še človeško interpretacijo. Potrebno bo torej »združiti najboljše, kar znamo ljudje, z najboljšim, kar znajo storiti računalniki, da izboljšamo zdravstveni sistem« (Dr. Lisa Latts).

6 LITERATURA

1. *The Surprising Things You Don't Know About Big Data* [Internet]. Adeptia; 2015 [cited 2021 Feb 6]. Available from: <https://adeptia.com/blog/surprising-things-you-dont-know-about-big-data>.
2. *Heads of medicines agencies - European medicines agency. HMA-EMA Joint big data taskforce summary report* [Internet]. European Medicines Agency; 2019 [cited 2021 Feb 6]. Available from:

- https://www.ema.europa.eu/en/documents/minutes/hma/ema-joint-task-force-big-data-summary-report_en.pdf
3. Fuller D, Buote R, Stanley K. A glossary for big data in population and public health: discussion and commentary on terminology and research methods. *J Epidemiol Community Health*. 2017;71(11):1113-7.
 4. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst*. 2014;2:3.
 5. Lee CH, Yoon HJ. Medical big data: promise and challenges. *Kidney Res Clin Pract*. 2017;36(1):3-11.
 6. De Mauro A, Greco M, Grimaldi M. A formal definition of Big Data based on its essential features. *Libr Rev*. 2016;65(3):122-35.
 7. Baker R. Big Data: A Survey Research Perspective. In: Biemer PP, editors. *The survey error in practice*. 1st ed. Wiley; 2017. p. 47-65.
 8. European Medicines Agency. Identifying opportunities for 'big data' in medicines development and regulatory science [Internet]. European Medicines Agency; 2017 [cited 2021 Feb 6]. Available from: https://www.ema.europa.eu/en/documents/report/report-workshop-identifying-opportunities-big-data-medicines-development-regulatory-science_en.pdf
 9. Callegaro M, Yang Y. The Role of Surveys in the Era of "Big Data". In: Vannette DL, Krosnick JA, editors. *The Palgrave Handbook of Survey Research*. 1st ed. Palgrave Macmillan; 2018. p. 175-92.
 10. Khoury MJ, Ioannidis JP. Medicine. Big data meets public health. *Science*. 2014;346(6213):1054-5.
 11. Covid-19 Sledilnik. Podatki o širjenju bolezni COVID-19 v Sloveniji [Internet]. [cited 2021 Feb 6]. Available from: <https://covid-19.sledilnik.org/sl/stats>.
 12. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*. 2021;589(7840):82-7.
 13. Burden AM. Pharmacoepidemiology and Big Data Analytics: Challenges and Opportunities when Moving towards Precision Medicine. *Chimia (Aarau)*. 2019;73(12):1012-7.
 14. Schneeweiss S. Improving therapeutic effectiveness and safety through big healthcare data. *Clin Pharmacol Ther*. 2016;99(3):262-5.
 15. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big Data In Health Care: Using Analytics To Identify And Manage High-Risk And High-Cost Patients. *Health Affair*. 2014;33(7):1123-31.
 16. O'Driscoll A, Daugelaite J, Sleator RD. 'Big data', Hadoop and cloud computing in genomics. *J Biomed Inform*. 2013;46(5):774-81.
 17. Ginsburg GS, Phillips KA. Precision Medicine: From Science To Value. *Health Aff (Millwood)*. 2018;37(5):694-701.
 18. Leyens L, Reumann M, Malats N, Brand A. Use of big data for drug development and for public and personal health and care. *Genet Epidemiol*. 2017;41(1):51-60.
 19. Brown N, Cambrozzi J, Cox PJ, Davies M, Dunbar J, Plumbley D, et al. Big Data in Drug Discovery. *Prog Med Chem*. 2018;57(1):277-356.
 20. Mittelstadt BD, Floridi L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Sci Eng Ethics*. 2016;22(2):303-41.
 21. Salerno J, Knoppers BM, Lee LM, Hlaing WM, Goodman KW. Ethics, big data and computing in epidemiology and public health. *Ann Epidemiol*. 2017;27(5):297-301.

