Trends in Population-Based Studies: Molecular and Digital Epidemiology (Review)

N.S. Denisov, Assistant, Research and Educational Laboratory "Live Laboratory
of Population-Based Studies";
E.M. Kamenskikh, Head of the Research and Educational Laboratory "Live Laboratory
of Population-Based Studies";
O.S. Fedorova, MD, DSc, Professor, Head of Department of Faculty Pediatrics with the Course
of Childhood Diseases, Faculty of General Medicine

Siberian State Medical University, 2 Moskovsky Trakt, Tomsk, 634050, Russia

The development of high-throughput technologies has sharply increased the opportunities to research the human body at the molecular, cellular, and organismal levels in the last decade. Rapid progress in biotechnology has caused a paradigm shift in population-based studies. Advances in modern biomedical sciences, including genomic, genome-wide, post-genomic research and bioinformatics, have contributed to the emergence of molecular epidemiology focused on the study of the personalized molecular mechanism of disease development and its extrapolation to the population level. The work of research teams at the intersection of information technology and medicine has become the basis for highlighting digital epidemiology, the important tools of which are machine learning, the ability to work with real world data, and accumulated big data.

The developed approaches accelerate the process of collecting and processing biomedical data, testing new scientific hypotheses. However, new methods are still in their infancy, they require testing of application under various conditions, as well as standardization. This review highlights the role of omics and digital technologies in population-based studies.

Key words: population-based studies in medicine; digital epidemiology; molecular epidemiology; omics technologies; information technologies.

Introduction

The global medical and demographic problems, those of population ageing, an increase in the prevalence of chronic non-communicable diseases, the pandemic of a new coronavirus infection, set new large-scale challenges for healthcare, where precision medicine becomes one of the tools for solving them. Initially demanded mainly in the diagnosis and treatment of oncological diseases, it is being introduced into all areas of medicine now.

Major research projects and campaigns are being initiated worldwide to develop and implement precision medicine strategies. Experts estimate the global precision medicine market to reach $87.7 billion by 2023. The leading scientific institutions are located in the USA, United Kingdom, France, and China. Since 2018, the number of publications in the field of precision medicine has amounted to about 16 thousand worldwide.

Molecular and digital epidemiology is one of the main tools of precision medicine.

Molecular epidemiology

Genomic research

Biological research has traditionally been carried out using reductionist approaches, partly due to

---

Corresponding author: Nikita S. Denisov, e-mail: denisov.ssmu@gmail.com

limitations in both the experimental power of the devices and the complexity of the analytical data evaluation processes. In the last decade, the development of high-throughput technologies has led to a sharp increase in the opportunities of studying the human body at the molecular, cellular, and organismal levels [1]. The rapid progress of biotechnology has led to a paradigm shift in genomic epidemiology, from linkage analysis to genome-wide association studies (GWAS) and the widespread use of next-generation sequencing (NGS). Technological developments have improved research design, enhanced our understanding of disease etiology, and led to numerous scientific discoveries [2].

In genomics, first-generation sequencing methods could sequence the human genome for $300,000; two decades later, next-generation methods can sequence the human genome in a few hours at a cost of $1000. Measurements of characteristics such as epigenome, transcriptome, proteome, etc. have undergone similar changes, which has allowed researchers to start studying pathologies using their characteristics at the molecular level rather than tissue one [3]. Therefore, both the study of individual organisms and the study of populations require calculative and statistical approaches to the data of various "omics", which consider metabolism in cells, tissues and organs as a whole, as an integrated system, rather than isolated separate processes.

The reductions in the cost of genome sequencing, combined with an increase in the computational power, have caused a strong revival of interest in the application of whole genome sequencing in public health [4]. Today, genomic epidemiology makes it possible to study the genomes of pathogens so as to have a better insight into the spread of infectious diseases among populations and quickly respond to the outbreaks of the diseases [5]. Together with philodynamics (a combination of epidemiology, evolution, and immunodynamics), genomic epidemiology is a rapidly developing field of science that addresses key issues related to epidemic preparedness and management in real time [6].

In the beginning, genomic data were used to study a variety of viruses, particularly, the influenza A virus and human immunodeficiency virus (HIV). The Ebola virus epidemic in West Africa (2013–2016) was the first major and large-scale challenge to study the virus genomes; that resulted in the discovery of their origin and causes for such a rapid spread of the epidemic and also allowed to detect subsequent sources of local outbreaks [7].

Genomic epidemiology has become a valuable source of information for scientists about the nature of the threats to public health such as Zika, Middle East Respiratory Syndrome (MERS), Ebola, and SARS-CoV-2 outbreaks [8]. These threats have required a variety of approaches including intensive genome sequencing to understand transmission dynamics during the acute phase of epidemics (Ebola virus in the Democratic Republic of the Congo) and broader genomic "surveillance" to detect a hidden increase in the prevalence (poliomyelitis) [9]. During the SARS-CoV-2 pandemic, many countries that had not previously used genomic data began to actively conduct such studies and rely on their results. Genomic technologies have made more than 2.5 million SARS-CoV-2 sequences known from over 185 countries [10], and due to the subsequent public interest in genomic epidemiology, new methodologies have been rapidly developed to fully utilize this dataset to fight against the pandemic.

The transmission of all infections occurs at different spatial scales, which depend on the pathogen, the nature of the host's movement, immunity, and other factors [11].

The impact of obtaining genomic data on the formation of public health is shown in Figure 1.

Genomic data can be used to characterize clinical cases of infection depending on location and time and track outbreaks at all spatial scales: from nosocomial infections to pandemics [12]. The analysis of the pathogen genomes in the context of other sequences obtained from the same outbreak, as well as their comparison with previously characterized variants, allow researchers to develop intervention strategies at the individual and population levels to minimize the burden of infectious diseases on the individual and society [13]. This comprehensive approach involving pathogen sequencing, analysis, and response is called
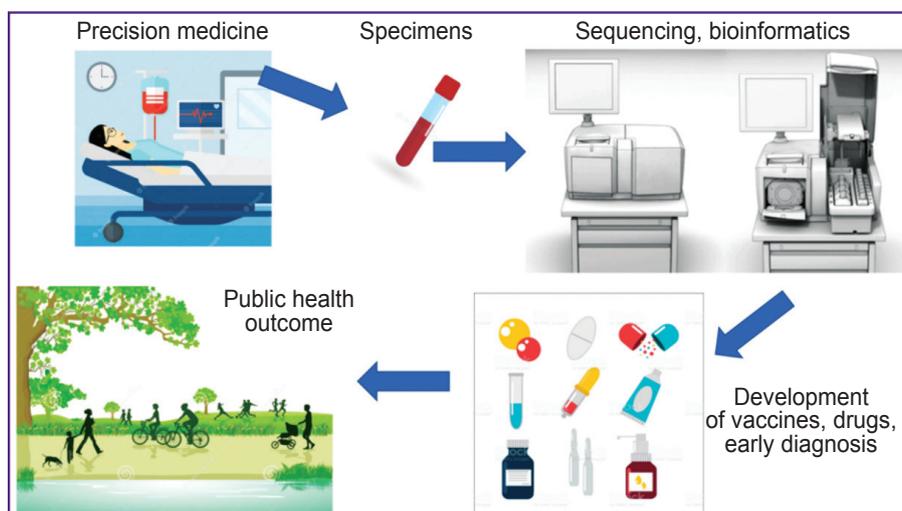
Figure 1. Pathogen sequencing during infectious disease outbreaks
Schematically presented relationship between obtaining genomic data and creating health-saving technologies

molecular epidemiology. In contrast to the development of individual-level treatment strategies that focus on the functional roles of host and/or pathogen mutations, the outbreak-scale genomic analysis uses pathogen mutations as markers of transmission events [14].

Genomic epidemiology studies the dynamics of outbreaks and the rapid evolution of pathogens that often accumulate mutations on the same scale as the spread of these pathogens.

NGS makes it possible to detect various types of genomic and epigenetic variations with high accuracy. Such sequencing allows researchers to directly study all these variations in person, increasing the chance of detecting mutations [15]. Although the use of NGS is still limited due to its high cost, the success of several recent projects demonstrates the great potential of this method in genomic epidemiology, especially in view of the sequencing cost decline.

With a sufficient sample size, appropriate metadata (such as location and date), and an appropriate statistical framework, pathogen genomes may assist in the identification of patterns in the spread of an epidemic with a small number of patients studied, allowing the development of precise targeted interventions compared to traditional methods and the use of demographic data [16]. In the nearest future, we will also be able to estimate the prevalence of chronic noncommunicable diseases using patient's pedigree data.

In 2011, the National Human Genome Research Institute (USA) published a review on genetic medicine, noting that the most effective way to improve human health is to understand normal biology (in this case, biology of the human genome) as a basis for studying the biology of diseases, which then becomes the basis for health promotion. To date, it is still difficult to fully determine the future prospects of genetic epidemiology for improving the public health [17].

When evaluating the contribution of genetic epidemiology to public health, it is equally important to understand that the etiology of diseases is complex and the genetic risk for developing pathology does not equate to genetic determinism [18]. The complex relationship between genetics and disease poses an ethical dilemma for practitioners regarding the correct interpretation of genetic test results. When performing genetic tests, it is possible to indirectly reveal the disorders that will not cause the development of the clinical disease manifestation [19]. An ethical question arises, should patients be aware of these incidental findings that may have a medical value?

Biomarkers

In the epidemiological study of diseases, metabolite concentrations are increasingly used as biomarkers that serve as indirect indicators of the rate of metabolic reactions. Though, the assessment of the rate of individual reactions can provide more accurate information about the ongoing changes directly in the organ [20].

Direct measurement of the rate of metabolic reactions *in situ* is currently impractical in large population studies since they are costly, technically complex, and require high-throughput equipment. This method is more successful when applied on a smaller scale, primarily through the use of non-invasive nuclear magnetic resonance spectroscopy (NMR spectroscopy) [21]. Metabolic pathway imaging techniques using hyperpolarized metabolites have shown promising results in the diagnosis and localization of tumors in patients with prostate cancer [22].

In a prospective clinical study involving 58 patients with chronic heart failure, the rate of adenosine triphosphate (ATP) synthesis was measured by studying the activity of cardiac creatine kinase *in situ* using the 31P NMR spectroscopy method [23].

ATP and creatine phosphate concentrations, as well as general clinical parameters, were used as predictors of chronic heart failure over an 8-year follow-up period. Excessive creatine kinase activity exceeded the significance of such parameters like patient's age, gender, and concentrations of other metabolites in predicting heart failure events and death, including hospitalization for heart failure and ventricular assist device insertion [24].

These results relate to a relatively small group of patients, but they add weight to the case for the development of biomarkers based on the rate of metabolic pathways and reactions in the study of disease.

Metabolism works as a continuously operating system of movement and transformation of molecules through reactions. Since the flow of metabolites is regularly redirected, metabolites are accumulated at various points or become depleted which results in a change in their concentration. The concentrations of metabolites reflect the effects of combined changes in the reaction rate, but do not give a direct idea of the dysfunctions of the processes themselves, for example, in pathology affecting enzymes, genes, and other molecular products derived from the human genome [24]. In this regard, a systematic assessment of the reaction rate on the scale required for epidemiology will be done by integrating metabolomic data with genomic, transcriptomic and/or proteomic information to determine enzymatic function.

Due to the ability to characterize diverse variants of endogenous and exogenous metabolites in biological specimens, metabolomic approaches have quickly been recognized as an important tool in public health studies [25]. The results show that the use of small volumes of blood, urine, feces, saliva, exhaled air condensate, cerebrospinal fluid, and biopsy for measuring the metabolome can provide information on possible mechanisms underlying the disease [26–30]. However, most of the existing evidence has come from case–control or crossover studies, which do not allow

for a clear temporal relationship between exposure, biomarkers, and disease.

Recently, the metabolic characterization of amniotic fluid, cord blood, and maternal/child urine or serum samples has been used to assess complex effects on the fetus and mother, and it may potentially be associated with developmental problems. Dried newborn blood spot used to identify metabolic biomarkers of future risk for cancer and other diseases have been proposed as a promising sample for metabolomic profiling [31–34].

The application of metabolomics for the study of disease risks, screening, and treatment efficacy has yielded promising initial results, although the field is still under development. These studies include ones on neurodegenerative diseases [35], type 2 diabetes [36], cancer [37], HIV, tuberculosis [38], malaria [39], and cardiovascular diseases [40]. The next important step in the application of metabolomics to study the etiology of diseases and early detection of pathologies will be longitudinal studies, which have already shown their effectiveness in creating biological models of the environmental impact on humans [41, 42].

Digital epidemiology

To conduct large multicenter epidemiological studies, digital technologies are actively used to facilitate the processes of work planning, data remote collection and entry control, as well as subsequent result presentation and reuse [43–46].

Though the epidemiology of chronic noncommunicable diseases in Russia is still lagging behind infectious diseases [47], there is a need to create and implement digital services for epidemiology of chronic noncommunicable diseases [48]. The need is owing to the increase of omics technologies' availability, the accumulation of the many years results of research, the need to compare the findings of similar studies, and the increased requirements for practical application and implementation of the results [47, 49].

Digital systems for clinical research

The basis for conducting research in the field of precision medicine is the formation of databases of clinical information annotated with the data on the collected biomaterials for each clinical case [50, 51]. This significantly expands resource opportunities for research at the intersection of clinical areas when new members of the research team are involved or in the case of a long-term work [52].

Coppola et al. [50] emphasized the importance of combining primary data with paraclinical information, including data from imaging studies, in a digital system. According to the authors, a service for visual data processing should have the options not only to display, but also to analyze data, which requires pre-processing

and data markup. The selection of areas with suspected lung infiltration according to computed tomography (CT) data or with pathological signal foci in the magnetic resonance imaging (MRI) pictures can be an example. Integration of genomic analysis into the data system contributes to the development of genomics and radiomics (radiomics is aimed at creating mathematical models and computer algorithms that, through the analysis of medical images, such as MRI or CT images, provide a finding about the pathophysiological features of tissues) [50, 53]. According to the research teams accumulating biomedical data, the imaging biobank data are to be used in accordance with the already-known standards until specific standards have been developed [50, 54]. Harmonization of processing will make it possible to combine data from multi-omics studies and visual materials for the integration of phenotypic and genotypic data [50, 55].

Over the past 10 years, many medical institutions have collected integrated databases (integrated data repositories, IDRs) [56], which are collected from electronic medical records [57]. Based on the accumulated data, not only scientific hypotheses are tested, but also a clinical decision support system is built [56]. Gagalova et al. [56] identified four models for the architecture of medical data collection and storage, in which data sources, the purpose of use, the availability of storage, etc. The purpose of this work was to initiate the development of guidelines on IDR creation in hospitals.

Online databases

Interactive monitoring systems have gained wide popularity [58]. Over the past 20 years, many services for monitoring infectious diseases have emerged [59, 60]. To monitor the situation with antibiotic resistance, many services have been created that are limited geographically as well as by described microorganisms and assessed metrics:

EARS-Net (https://atlas.ecdc.europa.eu/public/index.aspx);

CDDEP Resistance Map (https://resistancemap.cddep.org/index.php);

SGSS (https://sgss.phe.org.uk/Security/Register);

ATLAS (https://atlas-surveillance.com/#/login);

SMART (https://globalsmartsite.com/#/auth/login).

The free-access web application AMRmap (https://amrmap.ru/) [61] is a Russian development which displays data on antibiotic resistance obtained in multicenter clinical trials. The system has a section of genetic markers. Information in the database has been stored since 1997, access provided free of charge.

Since 2018, the University of Bristol's project EpiGraphDB [62] has been developing, which is a data-based analytical platform designed for the intellectual analysis of epidemiological indicators. The project is developing approaches to the

interpretation of causal relationships in the systematic automated analysis of many phenotypes using data from the array of bioinformatic resources. The university is also developing a software for statistical processing of omics studies, MR-Base being an example of it [63].

A large system of producing sequences of biological reactions in the body is presented in the WikiPathways system [64]. Currently, this system is being actively filled out with omics research data. The STRING database contains known and predicted protein–protein interactions [65].

Toom et al. [66] compared the results of an epidemiological study of headache in Estonia using an online questionnaire with the results of data research obtained during face-to-face visits of patients. The use of online questionnaires can significantly speed up the data collection process, increase population coverage, and reduce manual data entry errors. However, the authors noted that in the online survey, the majority of people did not have a headache, which greatly differed the sample of people who completed the online questionnaires from the sample of patients who came for face-to-face visits. This reduced the incidence of headache in the population. Also, more women, young people, married people, urban residents and people with a high level of education participated in the online survey. These characteristics of the sample are typical and should be considered as limiting in the case of studies using online questionnaires [67–69].

The integrated (online access, telephone, and paper mail) National Australian StepUp System for Dementia Research [70] is an interesting solution. In this system, patients with dementia and researchers of the diseases accompanied by cognitive deficits are registered in one of three convenient ways [70]. This allows accelerating the process of collecting data for research hypotheses and developing new approaches to combat dementia [71, 72]. The authors note that the continuous operation of the system went on after the start of the pandemic of a new coronavirus infection [70]. Over the two years of the platform operation, more than 1000 patients, 120 researchers have been registered, and more than 40 studies have been initiated [70].

For clinical trials, there are a number of free services which provide creating electronic individual registration cards, such as REDCap [73] or Ark [74]. The use of specialized services may be limited since access is provided to the organization after the conclusion of an agreement with the copyright holders and not directly to the researcher. However, the service ensures secure personal data storing without third-party access, unlike many open resources, including Google Forms [75]. In the future, research services will be used to create large databases on certain nosologies, diagnostic methods, or treatment. Services are constantly evolving, additional specialized analysis modules are created, for example, building a pedigree [74].

The pandemic of a new coronavirus infection caused an accelerated and forced introduction of digital technologies in all spheres of life, including all stages of research [76, 77]. Since the beginning of the pandemic in 2019, many national and international online monitoring systems have been developed [78]. The challenges for the fast-growing services are their weak integration with each other and the lack of centralized management, a difficulty in interpretation and practical application of data [79]. On the other hand, a limiting factor is the reluctance of patients to use digital questionnaires or remote methods of communication due to uncertainty about confidentiality in their use or unwillingness to become addicted to gadgets [80], which is especially common among older patients.

Open data

The annual increase in the accumulated data requires the introduction of new guidelines for the management of captured data. One of the most common standards for such work with data is FAIR (findability, accessibility, interoperability, and reusability) [81], which has become a fundamental requirement for open science [82, 83]. In their paper, Suhre et al. [84] emphasize the importance of data exchange for omics research, giving an example of a combination of GWAS and proteomic analysis. The authors consider the prospects for the creation of a database that will accumulate information about the genetic colocalization of genomic information and characteristics of the molecular phenotype of a disease (for example, gene expression and metabolomic characteristics) with clinical trial endpoints.

Real world data

Real world data in biomedical research refers to data captured from electronic medical records, medical registries, medical insurance companies, non-interventional clinical trials, and other sources in which information was obtained not under experimental conditions [85].

The HealthMap online system (https://www.healthmap.org/ru/) has been operating since 2006, accumulating data on disease outbreaks from open web resources [86]. In 2008, the web-based influenza surveillance system Influenzanet was launched [87, 88]. Limitations in the use of these data are their redundancy (repetitions), heterogeneity (different input formats), inconsistency (violation of the chronology of events). Chatzidimitriou et al. [89] created a database (n=20,463) on clinical cases of chronic lymphocytic leukemia (The ERIC CLL Database) filled with data from more than 90 centers and 31 countries. The authors consider the provision of standardization, integration of retrospective data, and assessment of the quality of input data to be necessary for the successful functioning of the distributed database [89].

Digital epidemiology as a separate field of knowledge

According to Salathé [90], digital epidemiology has become a separate area of scientific knowledge. Its purpose is to understand the patterns of disease development and the dynamics of the health status of the population, as well as to determine the causes of these patterns in order to find ways to prevent the development of diseases and promote health. The broadest definition of digital epidemiology is epidemiology that uses digital data. Though, the author then specifies that digital epidemiology operates the data that has not been collected with the main purpose of conducting epidemiological studies. Such data can include electronic medical records, information from insurance funds, city, regional, and federal health departments, as well as data from search engines, social networks, and mobile phones [90].

Google Flu Trends (GFT) has become one of the first known digital epidemiology services that uses search queries on acute respiratory symptoms for epidemiological analysis [91, 92]. A serious problem was that the collected data were owned by a private company, and the analysis algorithms used were unavailable even to national healthcare systems [90], and independent testing of the capabilities of this service for epidemiological studies showed a low efficiency in assessing the incidence of infectious diseases [93]. Unofficial Internet sources can be a valuable resource for epidemiological research, but the current trend towards protecting personal data and maintaining privacy is an important limiting factor. Salathé identifies two ways to the solution of this problem [90]:

creation of the monitoring systems by groups of scientists or professional communities, which will be more understandable and transparent for national healthcare, and that will increase the potential for their practical application;

greater involvement of the population in epidemiological studies. The rights to the data generated by individuals belong to the developers of the resource. A representative part of the population should be persuaded to share their personal health data with public health authorities for scientific research, the results of which can benefit society.

Roth et al. [94] have shown the formation of digital epidemiology (Figure 2).

According to the authors, machine learning methods based on the data from healthcare systems or social networks (Twitter), which help determine the prognosis for survival and complications, had already been developed by 2018.

It is important to note that the transformation of epidemiology leads to a change in its teaching principles [95]. Werler et al. [96] note that new curricula in epidemiology require the formation of causal thinking and the subsequent formation of a scientific hypothesis. Common mistakes made by young epidemiologists include estimation of one risk factor for one outcome, inaccurate formulation of research questions, and giving greater importance in research to epidemiological and statistical approaches compared to public importance.

Ethical issues

The development of high-precision medicine technologies entails the need to form new ethical standards [97]. Classical basic ethical principles are respect for patient's autonomy and privacy [98]. In this case, ethical requirements must ensure that individuals cannot be identified in open data portals for the exchange of scientific data. The ethics of precision public healthcare regulates the interaction between patients who have given voluntary informed consent to their attending doctor for the use of their clinical specimens in precision medicine research and the public decision-making process that drives public health activities. The development of a new hybrid ethical paradigm is possible only with the well-coordinated work of these process participants. Conducting omics studies allows obtaining detailed information about any subject. However, in order to plan disease control measures in a particular area or in a particular population, the following data indicating the demographic characteristics of an individual are important: geographical location, migration history, stay in prison, lifestyle and profession, etc. All these data are personal, they must not be subjected to
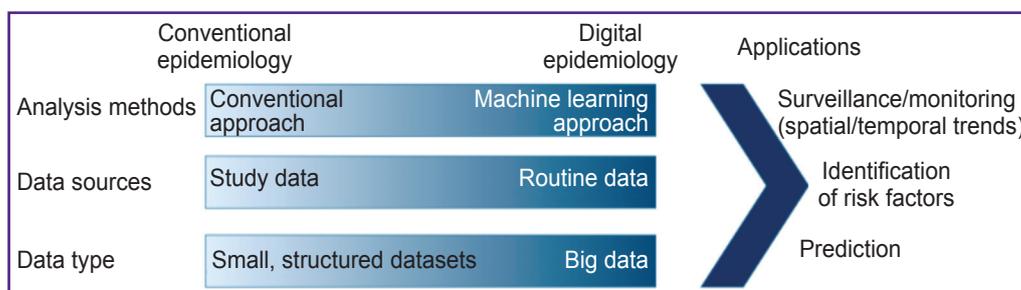
Figure 2. Transition from classical to digital epidemiology [94]
Formation of digital epidemiology as a field of knowledge

wide dissemination and increase the risks of discovering the identity of the subject.

In this regard, particular attention is paid to the way of presenting the obtained information. The ethics of precision medicine includes a public health ethic commitment to social justice and an emphasis on professional transparency and the trust built through it. The collected data should be transparent and aimed at improving the existing system and people's lives, and not stigmatizing social groups with high risk factors or relatively high incidence [97].

The development of electronic systems for capturing and storing data requires careful study of the risks to maintaining the security of the collected data [99]. New requirements for data management and professional confidentiality are emerging [98]. The speed, accuracy, and efficiency of big data processing offer great opportunities for public health, but entail a responsibility to adapt in a society that is committed to privacy, respect for human rights in matters of health, and social justice.

Sharma et al. [100] advocate for the development of legislation to maintain the confidentiality of personal data collected during scientific and clinical research, for auditing and implementation of independent oversight to assess the management of the risks related to the reuse of the data on research subjects. Solving this problem requires new approaches to working with patient data, taking into account an increased activity of scientific communication, creation of open repositories, exchange of primary research data, which is an integral part of large epidemiological studies. However, people are motivated to participate in study by pursuing their own interests, like the reputation of the organization with which they interact. Reuse of data by other organizations carries certain risks, which patients should be informed about before submitting voluntary informed consent to participate in a study.

FAIR-Health is a new paradigm of open science that has been developed in view of the peculiarities of biomedical research [101]. This paradigm is aimed at considering the information and biomaterials collected in research to be a single resource. It is this principle that, according to Holub et al. [101], will help ensure the reproducibility of studies and the subsequent integration of results.

Conclusion

Modern methods of population-based studies, including both omics technology data and the results of monitoring the conditions and behavior of patients over a long period of time, provide detailed data on subjects. At the moment, a search for methods of standardizing the collected data, their analysis and synthesis for further use is in progress. One of the major challenges to science is the integration of research results not only for rational storage, but also for the creation of dynamic digital models of subjects and processes.

The development of precision medicine technologies underlies the improvement of the quality of life and life expectancy of the population.

References

1. Diaz-Beltran L., Cano C., Wall D.P., Esteban F.J. Systems biology as a comparative approach to understand complex gene expression in neurological diseases. *Behav Sci (Basel)* 2013; 3(2): 253–272, https://doi.org/10.3390/bs3020253.

2. Wang Q., Lu Q., Zhao H. A review of study designs and statistical methods for genomic epidemiology studies using next generation sequencing. *Front Genet* 2015; 6: 149, https://doi.org/10.3389/fgene.2015.00149.

3. Pareek C.S., Smoczynski R., Tretyn A. Sequencing technologies and genome sequencing. *J Appl Genet* 2011; 52(4): 413–435, https://doi.org/10.1007/s13353-011-0057-x.

4. Hill V., Ruis C., Bajaj S., Pybus O.G., Kraemer M.U.G. Progress and challenges in virus genomic epidemiology. *Trends Parasitol* 2021; 37(12): 1038–1049, https://doi.org/10.1016/j.pt.2021.08.007.

5. Armstrong G.L., MacCannell D.R., Taylor J., Carleton H.A., Neuhaus E.B., Bradbury R.S., Posey J.E., Gwinn M. Pathogen genomics in public health. *N Engl J Med* 2019; 381(26): 2569–2580, https://doi.org/10.1056/nejmsr1813907.

6. Ladner J.T., Grubaugh N.D., Pybus O.G., Andersen K.G. Precision epidemiology for infectious disease control. *Nat Med* 2019; 25(2): 206–211, https://doi.org/10.1038/s41591-019-0345-2.

7. Holmes E.C., Dudas G., Rambaut A., Andersen K.G. The evolution of Ebola virus: insights from the 2013–2016 epidemic. *Nature* 2016; 538(7624): 193–200, https://doi.org/10.1038/nature19790.

8. Kulikowski C.A. Pandemics: historically slow "learning curve" leading to biomedical informatics and vaccine breakthroughs. *Yearb Med Inform* 2021; 30(1): 290–301, https://doi.org/10.1055/s-0041-1726482.

9. National Academies of Sciences, Engineering, and Medicine; Division on Earth and Life Studies; Board on Life Sciences; Health and Medicine Division; Board on Health Sciences Policy; Committee on Data Needs to Monitor the Evolution of SARS-CoV-2. *Genomic epidemiology data infrastructure needs for SARS-CoV-2: modernizing pandemic response strategies*. Washington (DC): National Academies Press (US); 2020. URL: https://www.ncbi.nlm.nih.gov/books/NBK562776/.

10. Liu T., Chen Z., Chen W., Chen X., Hosseini M., Yang Z., Li J., Ho D., Turay D., Gheorghe C.P., Jones W., Wang C. A benchmarking study of SARS-CoV-2 whole-genome sequencing protocols using COVID-19 patient samples.

iScience 2021; 24(8): 102892, https://doi.org/10.1016/j.isci.2021.102892.

11. Richard M., Knauf S., Lawrence P., Mather A.E., Munster V.J., Müller M.A., Smith D., Kuiken T. Factors determining human-to-human transmissibility of zoonotic pathogens via contact. *Curr Opin Virol* 2017; 22: 7–12, https://doi.org/10.1016/j.coviro.2016.11.004.

12. Gilchrist C.A., Turner S.D., Riley M.F., Petri W.A., Hewlett E.L. Whole-genome sequencing in outbreak analysis. *Clin Microbiol Rev* 2015; 28(3): 541–563, https://doi.org/10.1128/cmr.00075-13.

13. Grad Y.H., Lipsitch M. Epidemiologic data and pathogen genome sequences: a powerful synergy for public health. *Genome Biol* 2014; 15(11): 538, https://doi.org/10.1186/s13059-014-0538-4.

14. Wohl S., Schaffner S.F., Sabeti P.C. Genomic analysis of viral outbreaks. *Annu Rev Virol* 2016; 3(1): 173–195, https://doi.org/10.1146/annurev-virology-110615-035747.

15. Maljkovic Berry I., Melendrez M.C., Bishop-Lilly K.A., Rutvisuttinunt W., Pollett S., Talundzic E., Morton L., Jarman R.G. Next generation sequencing and bioinformatics methodologies for infectious disease research and public health: approaches, applications, and considerations for development of laboratory capacity. *J Infect Dis* 2020; 221(Suppl 3): S292–S307, https://doi.org/10.1093/infdis/jiz286.

16. Bandoy D.J.D.R., Weimer B.C. Analysis of SARS-CoV-2 genomic epidemiology reveals disease transmission coupled to variant emergence and allelic variation. *Sci Rep* 2021; 11(1): 7380, https://doi.org/10.1038/s41598-021-86265-4.

17. Gonzaga-Jauregui C., Lupski J.R., Gibbs R.A. Human genome sequencing in health and disease. *Annu Rev Med* 2012; 63: 35–61, https://doi.org/10.1146/annurev-med-051010-162644.

18. Cleeren E., Van der Heyden J., Brand A., Van Oyen H. Public health in the genomic era: will Public Health Genomics contribute to major changes in the prevention of common diseases? *Arch Public Health* 2011; 69(1): 8, https://doi.org/10.1186/0778-7367-69-8.

19. McGuire A.L., Beskow L.M. Informed consent in genomics and genetic research. *Annu Rev Genomics Hum Genet* 2010; 11: 361–381, https://doi.org/10.1146/annurev-genom-082509-141711.

20. Picó C., Serra F., Rodríguez A.M., Keijer J., Palou A. Biomarkers of nutrition and health: new tools for new approaches. *Nutrients* 2019; 11(5): 1092, https://doi.org/10.3390/nu11051092.

21. Fearnley L.G., Inouye M. Metabolomics in epidemiology: from metabolite concentrations to integrative reaction networks. *Int J Epidemiol* 2016; 45(5): 1319–1328, https://doi.org/10.1093/ije/dyw046.

22. Testa C., Pultrone C., Manners D.N., Schiavina R., Lodi R. Metabolic imaging in prostate cancer: where we are. *Front Oncol* 2016; 6: 225, https://doi.org/10.3389/fonc.2016.00225.

23. Luptak I., Sverdlov A.L., Panagia M., Qin F., Pimentel D.R., Croteau D., Siwik D.A., Ingwall J.S., Bachschmid M.M., Balschi J.A., Colucci W.S. Decreased ATP production and myocardial contractile reserve in metabolic heart disease. *J Mol Cell Cardiol* 2018; 116: 106–114, https://doi.org/10.1016/j.yjmcc.2018.01.017.

24. Metallo C.M., Heiden M.G.V. Understanding metabolic regulation and its influence on cell physiology. *Mol Cell* 2013; 49(3): 388–398, https://doi.org/10.1016/j.molcel.2013.01.018.

25. Johnson C.H., Ivanisevic J., Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* 2016; 17(7): 451–459, https://doi.org/10.1038/nrm.2016.25.

26. Bouatra S., Aziat F., Mandal R., Guo A.C., Wilson M.R., Knox C., Bjorndahl T.C., Krishnamurthy R., Saleem F., Liu P., Dame Z.T., Poelzer J., Huynh J., Yallou F.S., Psychogios N., Dong E., Bogumil R., Roehring C., Wishart D.S. The human urine metabolome. *PLoS One* 2013; 8(9): e73076, https://doi.org/10.1371/journal.pone.0073076.

27. Psychogios N., Hau D.D., Peng J., Guo A.C., Mandal R., Bouatra S., Sinelnikov I., Krishnamurthy R., Eisner R., Gautam B., Young N., Xia J., Knox C., Dong E., Huang P., Hollander Z., Pedersen T.L., Smith S.R., Bamforth F., Greiner R., McManus B., Newman J.W., Goodfriend T., Wishart D.S. The human serum metabolome. *PLoS One* 2011; 6(2): e16957, https://doi.org/10.1371/journal.pone.0016957.

28. Karu N., Deng L., Slae M., Guo A.C., Sajed T., Huynh H., Wine E., Wishart D.S. A review on human fecal metabolomics: methods, applications and the human fecal metabolome database. *Anal Chim Acta* 2018; 1030: 1–24, https://doi.org/10.1016/j.aca.2018.05.031.

29. Wishart D.S., Lewis M.J., Morrissey J.A., Flegel M.D., Jeroncic K., Xiong Y., Cheng D., Eisner R., Gautam B., Tzur D., Sawhney S., Bamforth F., Greiner R., Li L. The human cerebrospinal fluid metabolome. *J Chromatogr B* 2008; 871(2): 164–173, https://doi.org/10.1016/j.jchromb.2008.05.001.

30. Brown M.V., McDunn J.E., Gunst P.R., Smith E.M., Milburn M.V., Troyer D.A., Lawton K.A. Cancer detection and biopsy classification using concurrent histopathological and metabolomic analysis of core biopsies. *Genome Med* 2012; 4(4): 33, https://doi.org/10.1186/gm332.

31. Robinson O., Keski-Rahkonen P., Chatzi L., Kogevinas M., Nawrot T., Pizzi C., Plusquin M., Richiardi L., Robinot N., Sunyer J., Vermeulen R., Vrijheid M., Vineis P., Scalbert A., Chadeau-Hyam M. Cord blood metabolic signatures of birth weight: a population-based study. *J Proteome Res* 2018; 17(3): 1235–1247, https://doi.org/10.1021/acs.jproteome.7b00846.

32. Perng W., Rifas-Shiman S.L., McCulloch S., Chatzi L., Mantzoros C., Hivert M.F., Oken E. Associations of cord blood metabolites with perinatal characteristics, newborn anthropometry, and cord blood hormones in project viva. *Metabolism* 2017; 76: 11–22, https://doi.org/10.1016/j.metabol.2017.07.001.

33. Lau C.E., Siskos A.P., Maitre L., Robinson O., Athersuch T.J., Want E.J., Urquiza J., Casas M., Vafeiadi M., Roumeliotaki T., McEachan R.R.C., Azad R., Haug L.S., Meltzer H.M., Andrusaityte S., Petraviciene I., Grazuleviciene R., Thomsen C., Wright J., Slama R., Chatzi L., Vrijheid M., Keun H.C., Coen M. Determinants of the urinary and serum metabolome in children from six European populations. *BMC Med* 2018; 16(1): 202, https://doi.org/10.1186/s12916-018-1190-8.

34. Petrick L., Edmands W., Schiffman C., Grigoryan H., Perttula K., Yano Y., Dudoit S., Whitehead T., Metayer C., Rappaport S. An untargeted metabolomics method for archived newborn dried blood spots in epidemiologic studies. *Metabolomics* 2017; 13(3): 27, https://doi.org/10.1007/s11306-016-1153-z.

35. Roede J.R., Uppal K., Park Y., Lee K., Tran V., Walker D., Strobel F.H., Rhodes S.L., Ritz B., Jones D.P. Serum metabolomics of slow vs. rapid motor progression

Parkinson's disease: a pilot study. *PLoS One* 2013; 8(10): e77629, https://doi.org/10.1371/journal.pone.0077629.

36. Guasch-Ferré M., Hruby A., Toledo E., Clish C.B., Martínez-González M.A., Salas-Salvadó J., Hu F.B. Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis. *Diabetes Care* 2016; 39(5): 833–846, https://doi.org/10.2337/dc15-2251.

37. Wishart D.S., Mandal R., Stanislaus A., Ramirez-Gaona M. Cancer metabolomics and the human metabolome database. *Metabolites* 2016; 6(1): 10, https://doi.org/10.3390/metabo6010010.

38. Frediani J.K., Jones D.P., Tukvadze N., Uppal K., Sanikidze E., Kipiani M., Tran V.T., Hebbar G., Walker D.I., Kempker R.R., Kurani S.S., Colas R.A., Dalli J., Tangpricha V., Serhan C.N., Blumberg H.M., Ziegler T.R. Plasma metabolomics in human pulmonary tuberculosis disease: a pilot study. *PLoS One* 2014; 9(10): e108854, https://doi.org/10.1371/journal.pone.0108854.

39. Würtz P., Havulinna A.S., Soininen P., Tynkkynen T., Prieto-Merino D., Tillin T., Ghorbani A., Artati A., Wang Q., Tiainen M., Kangas A.J., Kettunen J., Kaikkonen J., Mikkilä V., Jula A., Kähönen M., Lehtimäki T., Lawlor D.A., Gaunt T.R., Hughes A.D., Sattar N., Illig T., Adamski J., Wang T.J., Perola M., Ripatti S., Vasan R.S., Raitakari O.T., Gerszten R.E., Casas J.P., Chaturvedi N., Ala-Korpela M., Salomaa V. Metabolite profiling and cardiovascular event risk. *Circulation* 2015; 131(9): 774–785, https://doi.org/10.1161/circulationaha.114.013116.

40. Ganna A., Salihovic S., Sundström J., Broeckling C.D., Hedman Å.K., Magnusson P.K.E., Pedersen N.L., Larsson A., Siegbahn A., Zilmer M., Prenni J., Arnlöv J., Lind L., Fall T., Ingelsson E. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *PLoS Genet* 2014; 10(12): e1004801, https://doi.org/10.1371/journal.pgen.1004801.

41. Rogers A.J., McGeachie M., Baron R.M., Gazourian L., Haspel J.A., Nakahira K., Fredenburgh L.E., Hunninghake G.M., Raby B.A., Matthay M.A., Otero R.M., Fowler V.G., Rivers E.P., Woods C.W., Kingsmore S., Langley R.J., Choi A.M. Metabolomic derangements are associated with mortality in critically ill adult patients. *PLoS One* 2014; 9(1): e87538, https://doi.org/10.1371/journal.pone.0087538.

42. Playdon M.C., Ziegler R.G., Sampson J.N., Stolzenberg-Solomon R., Thompson H.J., Irwin M.L., Mayne S.T., Hoover R.N., Moore S.C. Nutritional metabolomics and breast cancer risk in a prospective study. *Am J Clin Nutr* 2017; 106(2): 637–649, https://doi.org/10.3945/ajcn.116.150912.

43. Anfinogenova Y.J., Trubacheva I.A., Serebryakova V.N., Popov S.V. Emerging trends and challenges of population-based cardiology. *Sibirskij zhurnal klinicheskoj i jeksperimental'noj mediciny* 2020; 34(4): 24–38, https://doi.org/10.29001/2073-8552-2019-34-4-24-38.

44. Garg S., Williams N.L., Ip A., Dicker A.P. Clinical integration of digital solutions in health care: an overview of the current landscape of digital technologies in cancer care. *JCO Clin Cancer Inform* 2018; 2: 1–9, https://doi.org/10.1200/cci.17.00159.

45. Thyagarajan B., Nelson H.H., Poynter J.N., Prizment A.E., Roesler M.A., Cassidy E., Putnam S., Amos L., Hickle A., Reilly C., Spector L.G., Lazovich D. Field application of digital technologies for health assessment in the 10,000 families study. *Cancer Epidemiol Biomarkers Prev* 2020; 29(4): 744–751, https://doi.org/10.1158/1055-9965.epi-19-0858.

46. Coathup V., Teare H.J., Minari J., Yoshizawa G., Kaye J., Takahashi M.P., Kato K. Using digital technologies to engage with medical research: views of myotonic dystrophy patients in Japan. *BMC Med Ethics* 2016; 17(1): 51, https://doi.org/10.1186/s12910-016-0132-2.

47. Briko N.I. Theoretical generalizations in epidemiology: from history to the present. *Epidemiologia i vakcinoprofilaktika* 2018; 17(5): 5–16, https://doi.org/10.31631/2073-3046-2018-17-5-5-16.

48. Jagodzinski A., Johansen C., Koch-Gromus U., Aarabi G., Adam G., Anders S., Augustin M., der Kellen R.B., Beikler T., Behrendt C.A., Betz C.S., Bokemeyer C., Borof K., Briken P., Busch C.J., Büchel C., Brassen S., Debus E.S., Eggers L., Fiehler J., Gallinat J., Gellißen S., Gerloff C., Girdauskas E., Gosau M., Graefen M., Härter M., Harth V., Heidemann C., Heydecke G., Huber T.B., Hussein Y., Kampf M.O., von dem Knesebeck O., Konnopka A., König H.H., Kromer R., Kubisch C., Kühn S., Loges S., Löwe B., Lund G., Meyer C., Nagel L., Nienhaus A., Pantel K., Petersen E., Püschel K., Reichenspurner H., Sauter G., Scherer M., Scherschel K., Schiffner U., Schnabel R.B., Schulz H., Smeets R., Sokalskis V., Spitzer M.S., Terschüren C., Thederan I., Thoma T., Thomalla G., Waschki B., Wegscheider K., Wenzel J.P., Wiese S., Zyriax B.C., Zeller T., Blankenberg S. Rationale and design of the Hamburg city health study. *Eur J Epidemiol* 2020; 35(2): 169–181, https://doi.org/10.1007/s10654-019-00577-4.

49. Shin Y.K., Koskinen V., Kouvonen A., Kemppainen T., Olakivi A., Wrede S., Kemppainen L. Digital information technology use and transnational healthcare: a population-based study on older Russian-speaking migrants in Finland. *J Immigr Minor Health* 2022; 24(1): 125–135, https://doi.org/10.1007/s10903-021-01301-9.

50. Coppola L., Cianflone A., Grimaldi A.M., Incoronato M., Bevilacqua P., Messina F., Baselice S., Soricelli A., Mirabelli P., Salvatore M. Biobanking in health care: evolution and future directions. *J Transl Med* 2019; 17(1): 172, https://doi.org/10.1186/s12967-019-1922-3.

51. Griffon N., Pereira H., Djadi-Prat J., García M.T., Testoni S., Cariou M., Hilbey J., N'Dja A., Navarro G., Gentili N., Nanni O., Raineri M., Chatellier G., Gómez De La Camara A., Lewi M., Sundgren M., Daniel C., Garvey A., Todorovic M., Ammour N. Performances of a solution to semi-automatically fill eCRF with data from the electronic health record: protocol for a prospective individual participant data meta-analysis. *Stud Health Technol Inform* 2020; 270: 367–371, https://doi.org/10.3233/shti200184.

52. Rorie D.A., Flynn R.W.V., Grieve K., Doney A., Mackenzie I., MacDonald T.M., Rogers A. Electronic case report forms and electronic data capture within clinical trials and pharmacoepidemiology. *Br J Clin Pharmacol* 2017; 83(9): 1880–1895, https://doi.org/10.1111/bcp.13285.

53. Mayerhoefer M.E., Materka A., Langs G., Häggström I., Szczypiński P., Gibbs P., Cook G. Introduction to radiomics. *J Nucl Med* 2020; 61(4): 488–495, https://doi.org/10.2967/jnumed.118.222893.

54. Benelli M., Barucci A., Zoppetti N., Calusi S., Redapi L., Della Gala G., Piffer S., Bernardi L., Fusi F., Pallotta S. Comprehensive analysis of radiomic datasets by RadAR. *Cancer Res* 2020; 80(15): 3170–3174, https://doi.org/10.1158/0008-5472.can-20-0332.

55. Scapicchio C., Gabelloni M., Barucci A., Cioni D., Saba L., Neri E. A deep look into radiomics. *Radiol Med*

N.S. Denisov, E.M. Kamenskikh, O.S. Fedorova

2021; 126(10): 1296–1311, https://doi.org/10.1007/s11547-021-01389-x.

56. Gagalova K.K., Leon Elizalde M.A., Portales-Casamar E., Görges M. What you need to know before implementing a clinical research data warehouse: comparative review of integrated data repositories in health care institutions. *JMIR Form Res* 2020; 4(8): e17687, https://doi.org/10.2196/17687.

57. Huser V., Kahn M.G., Brown J.S., Gouripeddi R. Methods for examining data quality in healthcare integrated data repositories. *Pac Symp Biocomput* 2018; 23: 628–633.

58. Ku W.Y., Nfor O.N., Liu W.H., Tantoh D.M., Hsu S.Y., Wang L., Chou T.Y., Liaw Y.P. Online community collaborative map: a geospatial and data visualization tool for cancer data. *Medicine (Baltimore)* 2019; 98(20): e15521, https://doi.org/10.1097/md.0000000000015521.

59. Commons R.J., Thriemer K., Humphreys G., Suay I., Sibley C.H., Guerin P.J., Price R.N. The Vivax Surveyor: online mapping database for Plasmodium vivax clinical trials. *Int J Parasitol Drugs Drug Resist* 2017; 7(2): 181–190, https://doi.org/10.1016/j.ijpddr.2017.03.003.

60. Dong E., Du H., Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020; 20(5): 533–534, https://doi.org/10.1016/s1473-3099(20)30120-1.

61. Kuzmenkov A.Y., Trushin I.V., Vinogradova A.G., Avramenko A.A., Sukhorukova M.V., Malhotra-Kumar S., Dekhnich A.V., Edelstein M.V., Kozlov R.S. AMRmap: an interactive web platform for analysis of antimicrobial resistance surveillance data in Russia. *Front Microbiol* 2021; 12: 620002, https://doi.org/10.3389/fmicb.2021.620002.

62. Liu Y., Elsworth B., Erola P., Haberland V., Hemani G., Lyon M., Zheng J., Lloyd O., Vabistsevits M., Gaunt T.R. EpiGraphDB: a database and data mining platform for health data science. *Bioinformatics* 2021; 37(9): 1304–1311, https://doi.org/10.1093/bioinformatics/btaa961.

63. Hemani G., Zheng J., Elsworth B., Wade K.H., Haberland V., Baird D., Laurin C., Burgess S., Bowden J., Langdon R., Tan V.Y., Yarmolinsky J., Shihab H.A., Timpson N.J., Evans D.M., Relton C., Martin R.M., Davey Smith G., Gaunt T.R., Haycock P.C. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 2018; 7: e34408, https://doi.org/10.7554/eLife.34408.

64. Slenter D.N., Kutmon M., Hanspers K., Riutta A., Windsor J., Nunes N., Mélius J., Cirillo E., Coort S.L., Digles D., Ehrhart F., Giesbertz P., Kalafati M., Martens M., Miller R., Nishida K., Rieswijk L., Waagmeester A., Eijssen L.M.T., Evelo C.T., Pico A.R., Willighagen E.L. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res* 2018; 46(D1): D661–D667, https://doi.org/10.1093/nar/gkx1064.

65. Szklarczyk D., Morris J.H., Cook H., Kuhn M., Wyder S., Simonovic M., Santos A., Doncheva N.T., Roth A., Bork P., Jensen L.J., von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 2017; 45(D1): D362–D368, https://doi.org/10.1093/nar/gkw937.

66. Toom K., Raidvee A., Braschinsky M. The applicability of web-based solutions in headache epidemiology research. *J Headache Pain* 2020; 21(1): 60, https://doi.org/10.1186/s10194-020-01141-2.

67. Egger M.J., Lukacz E.S., Newhouse M., Wang J., Nygaard I. Web versus paper-based completion of the epidemiology of prolapse and incontinence questionnaire. *Female Pelvic Med Reconstr Surg* 2013; 19(1): 17–22, https://doi.org/10.1097/spv.0b013e31827bfd93.

68. Tassiopoulos K., Roberts-Toler C., Fichtenbaum C.J., Koletar S.L. Web-based data collection for older adults living with HIV in a clinical research setting: pilot observational study. *J Med Internet Res* 2020; 22(11): e18588, https://doi.org/10.2196/18588.

69. Braekman E., Charafeddine R., Demarest S., Drieskens S., Berete F., Gisle L., Van der Heyden J., Van Hal G. Comparing web-based versus face-to-face and paper-and-pencil questionnaire data collected through two Belgian health surveys. *Int J Public Health* 2020; 65(1): 5–16, https://doi.org/10.1007/s00038-019-01327-9.

70. Jeon Y.H., Shin M., Smith A., Beattie E., Brodaty H., Frost D., Hobbs A., Kotting P., Petrie G., Rossor M., Thompson J., Vickers J., Waters D. Early implementation and evaluation of StepUp for Dementia Research: an Australia-wide dementia research participation and public engagement platform. *Int J Environ Res Public Health* 2021; 18(21): 11353, https://doi.org/10.3390/ijerph182111353.

71. Langbaum J.B., Karlawish J., Roberts J.S., Wood E.M., Bradbury A., High N., Walsh T.L., Gordon D., Aggarwal R., Davis P., Stowell C., Trisko L., Langlois C.M., Reiman E.M., Tariot P.N. GeneMatch: a novel recruitment registry using at-home APOE genotyping to enhance referrals to Alzheimer's prevention studies. *Alzheimers Dement J Alzheimers Assoc* 2019; 15(4): 515–524, https://doi.org/10.1016/j.jalz.2018.12.007.

72. Weiner M.W., Nosheny R., Camacho M., Truran-Sacrey D., Mackin R.S., Flenniken D., Ulbricht A., Insel P., Finley S., Fockler J., Veitch D. The Brain Health Registry: an internet-based platform for recruitment, assessment, and longitudinal monitoring of participants for neuroscience studies. *Alzheimers Dement J Alzheimers Assoc* 2018; 14(8): 1063–1076, https://doi.org/10.1016/j.jalz.2018.02.021.

73. Harris P.A., Taylor R., Minor B.L., Elliott V., Fernandez M., O'Neal L., McLeod L., Delacqua G., Delacqua F., Kirby J., Duda S.N.; REDCap Consortium. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform* 2019; 95: 103208, https://doi.org/10.1016/j.jbi.2019.103208.

74. Ranaweera T., Makalic E., Hopper J.L., Bickerstaffe A. An open-source, integrated pedigree data management and visualization tool for genetic epidemiology. *Int J Epidemiol* 2018; 47(4): 1034–1039, https://doi.org/10.1093/ije/dyy049.

75. Djenno M., Insua G.M., Pho A. From paper to pixels: using Google Forms for collaboration and assessment. *Libr Hi Tech News* 2015; 32(4): 9–13, https://doi.org/10.1108/lhtn-12-2014-0105.

76. Wang X. Establishment of an internet-based epidemiological survey data collection customized system model. *Front Public Health* 2021; 9: 761031, https://doi.org/10.3389/fpubh.2021.761031.

77. Golinelli D., Boetto E., Carullo G., Nuzzolese A.G., Landini M.P., Fantini M.P. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. *J Med Internet Res* 2020; 22(11): e22280, https://doi.org/10.2196/22280.

78. Budd J., Miller B.S., Manning E.M., Lampos V., Zhuang M., Edelstein M., Rees G., Emery V.C., Stevens M.M., Keegan N., Short M.J., Pillay D., Manley E., Cox I.J., Heymann D., Johnson A.M., McKendry R.A. Digital technologies in the public-health response to COVID-19. *Nat Med* 2020; 26(8): 1183–1192, https://doi.org/10.1038/s41591-020-1011-4.

79. He Z., Zhang C.J.P., Huang J., Zhai J., Zhou S., Chiu J.W., Sheng J., Tsang W., Akinwunmi B.O., Ming W.K. A new era of epidemiology: digital epidemiology for investigating the COVID-19 outbreak in China. *J Med Internet Res* 2020; 22(9): e21685, https://doi.org/10.2196/21685.

80. Horst B.R., Sixsmith A., Simeonov D., Mihailidis A. Demographic and psychographic factors of social isolation during the COVID-19 pandemic: the importance of technology confidence. *Front Public Health* 2021; 9: 749515, https://doi.org/10.3389/fpubh.2021.749515.

81. Wilkinson M.D., Dumontier M., Jan Aalbersberg I., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.W., da Silva Santos L.B., Bourne P.E., Bouwman J., Brookes A.J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C.T., Finkers R., Gonzalez-Beltran A., Gray A.J.G., Groth P., Goble C., Grethe J.S., Heringa J., 't Hoen P.A.C., Hooft R., Kuhn T., Kok R., Kok J., Lusher S.J., Martone M.E., Mons A., Packer A.L., Persson B., Rocca-Serra P., Roos M., van Schaik R., Sansone S.A., Schultes E., Sengstag T., Slater T., Strawn G., Swertz M.A., Thompson M., van der Lei J., van Mulligen E., Velterop J., Waagmeester A., Wittenburg P., Wolstencroft K., Zhao J., Mons B. Addendum: the FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2019; 6(1): 6, https://doi.org/10.1038/s41597-019-0009-6.

82. Munafò M.R., Hollands G.J., Marteau T.M. Open science prevents mindless science. *BMJ* 2018; 363: k4309, https://doi.org/10.1136/bmj.k4309.

83. Sanjana N.E. Voices of the new generation: open science is good for science (and for you). *Nat Rev Mol Cell Biol* 2021; 22(11): 709, https://doi.org/10.1038/s41580-021-00414-1.

84. Suhre K., McCarthy M.I., Schwenk J.M. Genetics meets proteomics: perspectives for large population-based studies. *Nat Rev Genet* 2021; 22(1): 19–37, https://doi.org/10.1038/s41576-020-0268-2.

85. Bian J., Lyu T., Loiacono A., Viramontes T.M., Lipori G., Guo Y., Wu Y., Prosperi M., George T.J., Harle C.A., Shenkman E.A., Hogan W. Assessing the practice of data quality evaluation in a national clinical data research network through a systematic scoping review in the era of real-world data. *J Am Med Inform Assoc* 2020; 27(12): 1999–2010, https://doi.org/10.1093/jamia/ocaa245.

86. Brownstein J.S., Freifeld C.C., Reis B.Y., Mandl K.D. Surveillance Sans Frontières: internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med* 2008; 5(7): e151, https://doi.org/10.1371/journal.pmed.0050151.

87. Land-Zandstra A.M., van Beusekom M., Koppeschaar C., van den Broek J. Motivation and learning impact of Dutch flu-trackers. *J Sci Commun* 2016; 15(1): A04, https://doi.org/10.22323/2.15010204.

88. Paolotti D., Carnahan A., Colizza V., Eames K., Edmunds J., Gomes G., Koppeschaar C., Rehn M., Smallenburg R., Turbelin C., Van Noort S., Vespignani A. Web-based participatory surveillance of infectious diseases: the Influenzanet participatory surveillance experience. *Clin Microbiol Infect* 2014; 20(1): 17–21, https://doi.org/10.1111/1469-0691.12477.

89. Chatzidimitriou A., Minga E., Chatzikonstantinou T., Moreno C., Stamatopoulos K., Ghia P. Challenges and solutions for collecting and analyzing real world data: the Eric CLL database as an illustrative example. *Hemasphere* 2020; 4(5): e425, https://doi.org/10.1097/hs9.0000000000000425.

90. Salathé M. Digital epidemiology: what is it, and where is it going? *Life Sci Soc Policy* 2018; 14(1): 1, https://doi.org/10.1186/s40504-017-0065-7.

91. Ginsberg J., Mohebbi M.H., Patel R.S., Brammer L., Smolinski M.S., Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009; 457(7232): 1012–1014, https://doi.org/10.1038/nature07634.

92. Lippi G., Mattiuzzi C., Cervellin G. Is digital epidemiology the future of clinical epidemiology? *J Epidemiol Glob Health* 2019; 9(2): 146, https://doi.org/10.2991/jegh.k.190314.003.

93. Cervellin G., Comelli I., Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J Epidemiol Glob Health* 2017; 7(3): 185–189, https://doi.org/10.1016/j.jegh.2017.06.001.

94. Roth J.A., Battegay M., Juchler F., Vogt J.E., Widmer A.F. Introduction to machine learning in digital healthcare epidemiology. *Infect Control Hosp Epidemiol* 2018; 39(12): 1457–1462, https://doi.org/10.1017/ice.2018.265.

95. Frérot M., Lefebvre A., Aho S., Callier P., Astruc K., Aho Glélé L.S. What is epidemiology? Changing definitions of epidemiology 1978–2017. *PLoS One* 2018; 13(12): e0208442, https://doi.org/10.1371/journal.pone.0208442.

96. Werler M.M., Stuver S.O., Healey M.A., LaMorte W.W. The future of teaching epidemiology. *Am J Epidemiol* 2019; 188(5): 825–829, https://doi.org/10.1093/aje/kwz039.

97. Vallury K.D., Baird B., Miller E., Ward P. Going viral: researching safely on social media. *J Med Internet Res* 2021; 23(12): e29737, https://doi.org/10.2196/29737.

98. Juengst E.T., Van Rie A. Transparency, trust, and community welfare: towards a precision public health ethics framework for the genomics era. *Genome Med* 2020; 12(1): 98, https://doi.org/10.1186/s13073-020-00800-y.

99. Solomon D.H., Rudin R.S. Digital health technologies: opportunities and challenges in rheumatology. *Nat Rev Rheumatol* 2020; 16(9): 525–535, https://doi.org/10.1038/s41584-020-0461-x.

100. Sharma A., Nilsen T.B., Czerwinska K.P., Onitiu D., Brenna L., Johansen D., Johansen H.D. Up-to-the-minute privacy policies via gossips in participatory epidemiological studies. *Front Big Data* 2021; 4: 624424, https://doi.org/10.3389/fdata.2021.624424.

101. Holub P., Kohlmayer F., Prasser F., Mayrhofer M.T., Schlünder I., Martin G.M., Casati S., Koumakis L., Wutte A., Kozera Ł., Strapagiel D., Anton G., Zanetti G., Sezerman O.U., Mendy M., Valík D., Lavitrano M., Dagher G., Zatloukal K., van Ommen G.B., Litton J.E. Enhancing reuse of data and biological material in medical research: from FAIR to FAIR-Health. *Biopreserv Biobank* 2018; 16(2): 97–105, https://doi.org/10.1089/bio.2017.0110.

N.S. Denisov, E.M. Kamenskikh, O.S. Fedorova