P2P in Public Health: From Particles to Populations

Ozren Polašek, Kristina Fišter, Silvije Vuletić, Josipa Kern
Andrija Stampar School of Public Health, Medical School, University of Zagreb, Rockefellerova 4, 10000 Zagreb; opolasek@snz.hr; kfister@snz.hr; jkern@snz.hr; svuletic@snz.hr

Abstract. Despite numerous changes introduced by the new technologies, the role and the meaning of public health within health has remained stable and important. Public health is also expanding from the strictly population-focus to genetics and nano-technology. This means that the fundamentals of public health need to be supplemented by the additional expertise in areas such as medical informatics, bioinformatics, genetic epidemiology and public health genomics. The first step in integration is to develop the means of joining these ideas and experts in the solving of multidisciplinary problems, aiming to supplement public health with the emerging new technologies, ideas and concepts.

Keywords. public health, biomedical informatics, grid, nanotechnology, bioinformatics, genetic epidemiology

1. Introduction

Throughout the history, public health has served a very important role – thinking and acting for the benefits of the health of the entire population, rather than a single patient at one time. It is the discipline dealing with the disease prevention and health promotion through the organised efforts and informed choices of society, organisations, public and private, communities and individuals [1]. The main dissimilarity of public health and health care system is that public health is vital to all of us all the time, compared with (clinical) health care which we need only some of the time, after we or somebody else has suggested or set a disease diagnosis [2].

Despite strong medical background and intimate links of public health with both basic medical and clinical sciences, the main advances introduced by the public health actions were generally seen in the fields of sanitation and overall increase in the quality of living conditions, with only a handful of strictly medical actions that have substantially increased the health of the population, including for example vaccination. Public health has changed over time from the general discipline which was often merged together with the social medicine, hygiene or disease surveillance into a number of subspecialties and fields which now relatively independently cover various aspects of public health, including interventions in general public health and health promotion, epidemiology and surveillance, health systems research and public health informatics, to name only a few.

Recent advances in general science, including genetics and physics, has caused the public health paradigm to shift substantially and pay more attention to the phenomena which are hierarchically lower than the population. This mainly pertains to genetics and genetic epidemiology, which has had an enormous growth in the past decade and over time managed to accumulate numerous findings related to basic genetic structure, human variation and genetic background of disease. Some of these findings already are or are believed to be of major public health importance for the future. These include various fundamental discoveries, such as gene mapping attempts by the polymorphic markers, most commonly based on the single nucleotide polymorphisms (SNP). SNPs are defined as “change in which a single base in the DNA differs from the usual base at that position” [3]. Their main advantages are favourable cost-effectiveness and the possibility to detect diseases that are caused by the genes in which (or in whose close proximity) that SNP is physically located in our genomes. The list of diseases that have been associated with certain SNPs is long, for example sickle cell anaemia [4], or interim risk states such as obesity [5]. Sometimes an SNP can be just a normal genetic variation in the population. Detecting the SNP is the special problem. It demands for applying the special biotechnology methods, usually nano-methods [6]. Some particles measured on the nano scale could be applicable to diagnostics and therapy [7]. Nanoparticle is the lowest known level applicable in medicine and health care.
today. However, the question of their toxicity or potential harm should be investigated [8]. However, the main advantage (almost in the form of the “goodwill factor”) is expected in the future application of nanotechnologies, which indeed sound like a promising target, especially if these will be widely available and applicable to public health.

The overwhelming interest for nanotechnology application in medicine is seen in the number of published articles that are indexed on the PubMed; the use of the nanomedicine as a MeSH term suggests that the number of published articles related to that term has had an exponential increase ever since 2003, making it easily possible to show even greater advances in the future (Fig. 1).

Figure 1. The number of published studies indexed by the PubMed, with “nanomedicine” used as the MeSH keyword

2. Human genetics – background of health and disease

Although the traditional consideration of public health usually focuses on either a particular population or the entire human population, some areas of public health are focused on the lower hierarchical and organizational levels. Perhaps the most prominent of these areas is genetic epidemiology, which attempts to describe variation in human genetic structure and understand the genetic determination of human traits. By doing this, genetic epidemiology also serves as the science which explains previously unknown pathways and enables better and more detailed understanding of some basic physiologic mechanisms. The example for this type of insight is the recent discovery of the completely new uric acid transporter, which was previously believed to be the fructose transporter that was completely unrelated to uric acid transport [9]. The magnitude and importance of this has several facets. Firstly, the main and the most prevalent type of studies in the modern genetic epidemiology is the genome-wide association study, which is by its nature a hypothesis-free investigation (as it focuses on the association of a large number of genetic markers with the given trait or diseases, without any prior knowledge on the genetic background of the selected trait). While the sceptics may question the usability of this method, it has proved as highly informative and effective in the discovery of the genetic background of human diseases. Secondly, the magnitude of this type of research has grown immensely, with relatively easily and cheaply available genome-scans with 500,000 or even one million of single-nucleotide markers. This large abundance of information has two main shortcomings. The first one is the need to ensure statistical corrections due to the large number of statistical tests that are being performed, all in the situation when the classic statistical power analyses do not provide sufficient information. Secondly, the large amount of information provided by the contemporary genotyping efforts presents a challenge for even the most basic data operations, with easily several millions of mathematical operations needed to perform even the most basic descriptive analysis. This is the area where information and communication technology (ICT), especially grid technology, has a central role of storing, managing and querying such data. Special tools have been developed, and are constantly being developed which enable more and more precise and targeted analyses, all aiming to describe variation in the genetic material among humans, and then try to use it as the causal or modifying effects of human diseases. This area is well developed, with rather frequent changes of the means: from microsatellite markers some five years ago, to contemporary short nucleotide polymorphisms, to the analysis of the rare variants and copy number variation in the future.

The main problem of the research of the commonest contemporary killer – cardiovascular diseases (and non-communicable diseases in the broad sense) is related to the complexity of their clinical appearance and mechanisms from which they originate. This is related to the disease phenotypes, such as e.g. myocardial infarction, which are as a rule the end-result of a large number of potential factors which can act either as protective agents (e.g. favourable diet,
physical activity, etc.) or unfavourable agents (smoking, fatty diet, etc.). Then we have to add various possible interactions between genes, genes and environment and the ubiquitous random effects (where both true random effects and those that we don’t know how to explain are often lumped together), and the complexity of the research area increases rapidly, and inversely, the prospects of easy discoveries decrease.

Once the potential genetic marker has been identified, several steps are needed. The first one is to ensure that the result was not false positive or false negative. This is most commonly performed by the use of one or several replication populations and the advancement of the research only if the replication holds. Next, another discipline enters the research area – bioinformatics. Bioinformatics attempts to describe and understand the DNA structure (or other biological molecules such as RNA and proteins) and predict the function of the given sequence or a segment. This is based primarily on the homology with the known mouse, yeast or fruit fly DNA sequences, where genetic manipulation has enabled much faster and easier understanding of basic biological and pathological processes. But, the true story is again much more complicated than this. Firstly, at the moment most of research on DNA is done on the markers, highly polymorphic sections of the genome. Step forward is the complete sequencing, in which not only those highly polymorphic sites will be available for the analysis, but the entire human genome, with 3*10^9 base pairs. Such vast quantity of data is currently a serious problem for the analysis, especially if we take into account possible complexity and interactions between various genes, further aggravated by the possible gene-environment interactions.

3. Information – background of evidence based health care

One of the best publicly available data and knowledge integration web services is the National Center for Biotechnology Information, which is a service of the U. S. National Library of Medicine and the National Institutes of Health (http://www.ncbi.nlm.nih.gov). Until recently, the most used section of this website was PubMed, a database for search of the Medline, the most popular medical research indexing database. PubMed currently covers over 18 million citations from Medline and some other associated sources (sources that might be of interest for medicine are occasionally added from other journals, including veterinary medicine, research methodology, psychology, etc.).

However, the same website offers a number of other sources, including extensive information on known and expected proteins, single-nucleotide polymorphisms and genes, as well as extensive information on known diseases associated with these markers, genes and proteins. The combination of these resources is very important in modern human genetic, due to the extreme complexity of the human genome. Currently, human genome is believed to comprise approximately 20,000-25,000 genes which govern all known variation in humans, cause and maintain our individuality, define the way in which our bodies are built and the ways in which our bodies respond to pathogenic and harmful effects. Even the simple extrapolation from these figures suggests that 25,000 distinctive genes are largely insufficient to govern all structure, mechanisms and pathology in the body [10]. Several additional layers of complexity are in place. Firstly, DNA holds the basic information. The DNA is then translated into the RNA (ribonucleic acid), which is used as the basis for protein creation. But then, proteins can be modified by the alternative splicing and glycosylation, they can have different actions in different tissues and in different organs. Thus, the amount of complexity increases immensely, allowing that only as much as 25,000 genes may maintain all variation but also suggesting that the amount of complexity is much greater that initially believed.

Another resource that deserves a mention is the http://www.genecards.org website, a service provided for free to academic non-profit institutions, set up and run by the XenneX, Inc. The website contains several items, including the GeneCards section, which enables users to find a specific gene of interest and link the gene with published articles in PubMed, thus allowing better and more thorough browsing and integrating genetics with the clinical and applied medicine. It also offers a number of very important resources, such as the summary for the selected gene from the main genetic epidemiology and bioinformatics resources, physical map for the gene and chromosome section where the gene has been described, known synonyms, expression profiles for most important tissues, ontology tools, list of available single-nucleotide markers associated with the
selected genes and information on the manufacturer of the genetic chips used in genotypedation, and lastly, links to known (human) diseases which can easily be summarized and analysed. Thus, the website provides a great integrated source of information for researchers who do not need to be highly involved in genetics, but are only interested in the clinical and applied possibilities of the selected marker or gene being used in clinical or public health medicine.

There are several attempts to find out better possibilities for searching for information on health and disease, methods and techniques. One of such results is Building an Index of Nanomedical Resources [11].

Overall, there seems to be a growing amount of information on sub-cellular level in public health, which is clearly seen in the creation of the journal “Public Health Genomics”. The journal focuses on the translation of genome-based knowledge and technology into public policy, disease prevention and improvement of population health; it is a peer-reviewed journal in a bimonthly forum featuring conceptual and original research articles, editorials, reviews, short communications, country reports, case studies, viewpoints, and other features about the nature of public health genomics and related concepts. The aim of 'Public Health Genomics' is to facilitate a wide dialogue among researchers, practitioners, policy-makers and community members [12].

4. Individual information

Being active and taking care of own health is a recent trend in every persons’ health care. Electronic health record (EHR) is the basis but (electronic) personal health record (PHR) seems to have become the even more interesting as it manages to attract the interest of individual for his or her own health rather than relying solely of the physician for recording and managing health data [13]. The fact that the content of PHR and what issues surrounding it is underdeveloped suggests that this is one of the areas where further advances are to be expected [14].

Medical informatics therefore has an important role of merging and linking between the micro- and macro-level, all in an attempt to create and manage individual health record and the creation of the basic premises needed for the “personalized medicine” concept, in which the treatment options and health related advices are guided by the individual characteristics, mainly her or his own genetic background that makes that person more or less prone to good therapeutic advances, but also serves as the source of information in which, based on the same set of genetic data the probability of side-effects is reduced.

In an attempt to cover all the facets of health and disease, from gene level to clinic, the EHR (and possibly PHR) should be developed, standardized and certificated. The functionality, interoperability and security are requirements sine qua non.

5. Integration in public health

In the attempt to intervene and to achieve better health of the population, public health should be familiar with potential and risks of population under study. Both potential and risks have genes and genetic traits in their background. The other side of potential and risks is in environment where they should be realized, e.g. ecosystem, lifestyle of population, and health service organization. Culture, tradition and social environment could also affect health. Public health should take care of all these facets, and insist on information system development which includes such data. Martin-Sanchez and Maojo suggest integration of genomics data into health information system and creation of new directions in public health [15,16]. It is obvious that this task strongly demands grid technology, enough capacities of ICT and human resources working together.

6. Initiatives

It is now clear that the main contemporary problem in public health and in the whole health area is to integrate the available data and methods which can all be used in public health and health area as the whole. One of such attempts to bring together various aspects of information in medicine is ACTION-Grid, a project on biomedical informatics (BMI), grid technologies and nanoinformatics. The project is funded by the European Commission and it brings together a consortium of European researchers along with researchers in Latin America, the Western Balkans and North Africa. The project will foster training and mobility in the sphere of biomedical informatics, grid- and nanotechnology methods and services. One of the project goals is also to develop White Paper
in collaboration with a panel of recognized experts. This document will be delivered to the EC to establish a future agenda covering the grid/nano/bio/medical informatics areas and to develop new research plans in regions included in the project (EU, Latin America, Western Balkans, North Africa). Yet, probably the most influential and the most important element of this project is the capacity building, by linking various professionals into the common network that can share ideas, knowledge and most importantly, various tasks in achieving the common goal.

7. Acknowledgements

This study was supported by the ACTION-Grid project, funded by the European Commission under the 7th framework project (Project Reference: 224176).

8. References


