BIG DATA FOR HEALTH SERVICE RESEARCH: BALANCING POTENTIALS AND CHALLENGES

By: Anna-Theresa Renner, Julia Bobek and Herwig Ostermann

Summary: Big data, and digitised information in general, is of high importance and already widely used in most sectors and research fields, including health. Purposeful application of health data can contribute to better population health and more efficient health service provision. Nevertheless, precautions need to be taken, as individual health data is highly sensitive and misuse can have significant negative effects on the individual (e.g. on the labour market). This article explores the potentials and pitfalls of using big data in health service research. Furthermore, it highlights the necessity for governance of the interests of different stakeholders in accessing health data.

Keywords: Big Data, Health Service Research, Health Data, Data Governance

Introduction
Digitisation of everyday life and technological advancements in the storage of collected data (server sizes) have led to the increasing relevance of data for research and business. As the processing power of conventional computers has steadily increased and the public mind-set has turned towards data driven information over recent decades, the term “big data” is frequently used in scientific and non-scientific discussions. Even though there is no single definition of big data, the term usually refers to very large amounts of data that are routinely or automatically collected and stored. Data can be structured or unstructured (e.g. pictures) and can be mined for information, whereas the insights of conventional inductive statistical inquiries are fairly limited for big data.

Quite often big data is defined by three characteristics, known as the “3Vs”:
- volume
- velocity
- variety
It has been further suggested to add value and veracity as fourth and fifth “Vs”. Other authors have proposed to define big data with respect to the sample population which equals the whole basic population. Although there is currently no definition of big data in health, the importance of the topic has been recognised by the European Commission and its Directorate General for Health and Food Safety (DG SANTE), which is currently developing policy recommendations for big data in public health, telemedicine and health care.

Potentials and pitfalls of big data
Health data are collected at different service levels of the health care sector (e.g. in hospitals, primary care or...
pharmacies), but also outside the health care sector (e.g. via mHealth-Apps or social media), and for different purposes, such as reimbursement and insurance claims, or for epidemiological reasons (e.g. registries). Other (secondary) uses of these data sources can be utilised for further applications such as health economic, health system or health service research, and of course clinical research (e.g. with genomic data). Data content ranges from genomics data to socioeconomic data, including, among others, data on pharmaceuticals, treatment processes, insurance claims, telemedicine, and on wellbeing and behaviours.

The possibilities of analysing these data are numerous, and possible research questions that can be answered increase even more when linking different datasets. Moreover, the costs of using big data for research are relatively low compared to data collected in clinical trials, but yield similar robust results due to the sheer volume. Utilising this potential of big data can benefit the single patient when used for research on effectiveness, quality, and safety of treatments and prevention, but also the whole population for example by using it for infectious disease monitoring. Furthermore, accessible big data facilitates comparative effectiveness research which will ultimately lead to cost-containment and more effective distribution of resources in the health care sector or the whole economy.

There are also some pitfalls related to big data in health that must not be neglected. These are mainly related to the fact that health data are not only individual-level data, but also highly sensitive, as misuse can negatively affect the individual, for example on the labour market or with regard to insurance payments. This is the reason why health data cannot be treated in the same way as data from other areas of life, but need special regulations. The European Parliament and the Council have recognised this fact in their recently ratified “General Data Protection Regulation” (GDPR), where health data are mentioned as one of the “special categories of personal data” [6, Article 9]. The GDPR allows for derogation from a prohibition on processing these special categories of personal data, only if its purpose is in the interest of the public, which includes to “ensure the quality and cost-effectiveness of the procedures […] in the health insurance system, or for […] scientific or historical research purposes or statistical purposes”. This means that big data for health service research is in principle allowed, but only when its benefit for the public is substantiated.

Another pitfall that is related to data protection is the secondary use of data. Quite often data are collected for a specific purpose (e.g. a clinical study) to which the patient or any other data subject (e.g. a health care provider) has consented to. In some countries, use of these data for any other purposes, including research, needs further approval of the data subject. This presents a severe barrier for scientific research. In some other countries, access to health data is permitted for research if it is done in the public interest, and the individual-level data are anonymised or pseudonymised. Under the GDPR the processing of health data without consent of the data subject is possible under the condition that the rights and freedoms of natural persons are protected by suitable and specific measures. The GDPR is generally seen as a step in the right direction to align European national legislation, but critics have raised the concern that there is too much room for interpretation on how it should be implemented in the EU Member States. Even though some EU Member States already have stricter data and privacy protection laws in place, the implementation of the GDPR, especially the appointment of a Data Protection Officer in each country, is feared to increase administrative burden and to require a high level of human resources.

Another important issue that could mitigate the benefits of big data in health service research is the quality of the data and the data analysis. To derive valid conclusions from quantitative analyses, researchers need to be aware of the quality of their analysed data. Accuracy, completeness, consistency, reliability, timeliness, and validity are frequently named as indicators of data quality. Especially the quality of data from mHealth apps is often unclear, but highly relevant when linking these data with e.g. routinely collected health care records. To tackle this issue, the European Commission has set up a working group that will develop guidelines for assessing the quality of data collected via health apps.

Other data sources and datasets comprise some intricacies as well. For example, diagnostic data that are collected in hospitals for reimbursement purposes, are generally regarded as high quality, but might be prone to up-coding, or a bias stemming from different coding routines in different hospitals. Being aware of these possible biases and using statistical methods to control for them is essential for health service research to produce robust results that eventually lead to more informed and evidence-based political decision-making.

Besides shortcomings in data quality, the quality of the data analysis is also key to the valuable utilisation of big data. In analysing big data, researchers should be aware that the probability of spurious correlations rise with the size of the available datasets. It is therefore imperative for the analysis of big data for health service research that analytical skills are paired with knowledge of the field. To exhaust the potentials of big data in health, researchers have to be able to identify, within the abundance of data, what information is crucial to answer a specific and relevant research question.

Using big data for research

Several projects in various EU Member States aim to facilitate health service research by linking relevant datasets. The Austrian project DEXHELPP (www.dexhelpp.at), which is co-funded by two ministries and the city of Vienna, uses existing health care data to develop methods, models and technologies for supporting decisions in health policy and planning. The project is coordinated by the Vienna University of Technology and carried out together with private and public partners, such as the Main Association of the Austrian Social Security Institutions (“Hauptverband der österreichischen Sozialversicherungsträger”) and the Austrian Public Health Institute (“Gesundheit Österreich GmbH”). By developing methods for linking different datasets, analyses of the current status but also models for forecasting and for comparative evaluations can be carried out.
One of the main achievements of this project so far has been the development of a secure research server for all project partners, where highly heterogeneous datasets can be safely stored and analysed. This server is the basis for other research areas of DEXHELPP, such as estimating the burden of disease with computer simulation models or the comparison of different health care interventions and payment systems. Many other European countries have implemented similar projects (e.g. UK Clinical Practice Research Datalink, Italian ARNO Observatory, Swedish ICT eHealth). Even though, the fields of application of these projects vary, the common objective is to make data available for research in the public interest.

The aforementioned research projects usually include data from electronic patient records or health records (EPR/EHR) and electronic prescription systems. The aim of such eHealth systems is to improve patient care pathways by enabling a secure exchange of the collected patient level data between health care providers. In Austria, the electronic health care record (“Elektronische Gesundheitsakte” – ELGA) (www.elga.gv.at), which includes an eMedication application, is currently being piloted in several regions. Other European countries are further ahead in the implementation of eHealth structures, such as the Netherlands (AORTA), Denmark (Shared Care Platform) and Estonia (E-Estonia national identity scheme).

There is currently no common understanding or guideline at the European level on which applications should be incorporated in a national eHealth structure or what the content of an EPR/EHR should be. However, efforts on this issue have been made by the European Union and its Member States, by co-funding the European Patient Smart Open Services (ePSOS) project (www.epsos.eu), with the objective of improving the interoperability of eHealth systems to facilitate cross-border health care in Europe. ePSOS focused on technical and semantic aspects, but also on legal and organisational frameworks, and developed recommendations for supporting further developments in cross-border interoperability. Making cross-border EPR/EHR usable for research was not a primary goal of the ePSOS. Nevertheless, aligning national eHealth structures, or at least defining a minimum level of technical and content-wise standardisation, will not only improve cross-border health care, but also enable cross-country comparisons for health service research.

Goverance of stakeholder interests

Reservations against the widespread use of big data, especially of big health data, should be taken seriously, especially when coming from the data subjects (i.e. patients or health care providers). Communicating the potential benefits of big data to citizens and stakeholders will be crucial, and has to be done in a measured way. It will not be enough to highlight the benefits, but it must be made clear under which circumstances the analysis of big data sets has an advantage over other methods of evidence generation. Furthermore, the fears of the data subjects regarding privacy need to be addressed openly, which includes informing them about existing legal frameworks, as well as other data protection policies to reduce possible data breaches or data abuse to a minimum.

Big data not only offers potential opportunities for individuals and public health, but is also a big business opportunity for companies in the health care sector. European industrial stakeholders have raised concerns that the relatively high data protection standards in the EU compared to other parts of the world, might shift business opportunities to countries outside the EU. Contrary to this fear, the European Commission hopes to attract business by increasing the trust of its citizens, which in turn enables companies to establish sustainable relationships with their clients. Moreover, the EU rules on data protection and privacy apply to all companies, including those from non-EU countries, which operate in an EU Member State. Whether these efforts will yield the expected results, or whether business opportunities will accelerate elsewhere, remains to be seen. Independent of these developments, business considerations, unless they are in the interest of the general public, should not compromise the privacy rights of citizens.

For health service research it is crucial that the process of accessing data (and big data in particular) for research purposes is transparent, and equal for all researchers. Therefore, data governance is a key issue in utilising the full potential of big data analysis. Data governance includes clear guidelines on what data can be used, in what form (pseudonymised, level of aggregation etc.) and by whom. This not only encompasses (public) health researchers but also state institutions in their role of planning and organising (public) health service provision. This way, big data governance can substantially contribute to accountability, not only of individual health care providers, but also of the state as a regulator for the provision of health and social services, hence, shaping a more equal relationship between the state and its citizens. Therefore, the value of big data for health service research are not reflected in the sheer amount of available and accessible data, but in the sensible use of these data to generate high level evidence that can be used for (better) policy making targeted at the welfare of the population.

References