American Journal of Biodiversity
https://biojournals.us/index.php/AJB
ISSN: 2997-3600

Harnessing Health Statistics for Predictive Analytics: Transforming Healthcare Outcomes and Personalized Medicine

Ali Malik Faleh Abdul Rahim
Al-Mustansiriya University College of Science Department of Physics

Fatima Saad Hamid Kalil, Abeer Mohamed Mahmoud
Middle Technical University, Institute of Management, Rusafa Health Statistics Department

Huda Ghazwan Ali jasim
Middle Technical University / Institute of Management Rasafah Section Health Statistics

Wiaam Adham Hashem Mohamed
Middle Technical University Institute of Management/Rusafa Department of Health Statistics Techniques

Received: 2024, 15, Dec
Accepted: 2025, 21, Jan
Published: 2025, 28, Feb

Annotation: Technology has enabled us to track, capture, and analyze more data than ever before—and this data is growing at an exponential rate. Data in the healthcare domain can be used to transform healthcare outcomes to deliver personalized medicine. This is due to the appearance of connected health infrastructure, in which a variety of medical devices comfortably coexist with traditional computing and storage devices such that healthcare services can be provided at any time and in any place. A central component of connected health is the individual and doctor-centric approach of telemedicine services that can monitor and analyze a person's medical condition without the need of physically entering a doctor's office. Long-term monitoring services create the need for distributed approaches that would allow remote analysis of health data collected from the various monitoring devices. Thus, the collection and analysis of health statistics is becoming faster and easier, and it is expanding rapidly. Harnessing health statistics provides the opportunity to reduce healthcare costs and delivers treatment more effectively. Moreover, it has the potential to enhance personal treatment, as well as enable early diagnosis which can be crucial for the treatment of severe diseases.

The increasing interest in analyzing health statistics creates a new range of challenges. One of the key challenges is the privacy of health statistics. Patient's health statistics can represent sensitive information, and therefore they must be stored and processed securely, not only to protect patient privacy as demanded by the law, but in addition to protect the commercial interests of healthcare providers. Healthcare providers store and analyze patient health statistics, and represent a source of internal threats and hazards. On the other hand, wearable sensors became popular among patients. Off-the-shelf mobile devices have integrated sensors that can monitor a set of health statistics. The resultant data has obvious privacy concerns when being collected and analyzed by third party services that can run on the same device.

1. Introduction to Health Statistics and Predictive Analytics

Health statistics, in particular electronic health records, and predictive analytics are powerful tools aimed at fostering the improvement of healthcare outcomes. More informed decision making will help achieve higher levels of well-being, allowing the adoption of proactive health improvement initiatives. Patient neural network allows multiple solutions determination from a given input dataset. The temporal dimension is not considered, and each dataset unique finding may not provide transportation to other datasets with equal attributes. However, despite these limitations, one main of the methods is the well-performing prediction model that can assist care providers in improving patient diagnosis. Finally, the epidemic situation has been deeply influenced by response, which has some future threats. Predictive analytics and machine learning are today playing an ever-important role in treatment plan formulations. It is being outsourced to complex systems leveraging AI. This ensures improvement on the accuracy of the early detection model, potentially reducing costs associated with possible future invasive interventions. With the advent of next-generation phenotyping, treatment plans will be geared for higher levels of success, paving the way towards levels of precision. In this context, a synthesis of health statistics and predictive analytics are pivotal for driving future development-oriented preventative treatments.

1.1. Overview of Health Statistics

At the heart of many challenges of the healthcare industry today is how to leverage health data to analyse, rationalize diverse outcomes, and effectively personalize treatment. Research that translates and transforms health data into structured information related to the treatment of patients can enable better doctoral workflow and improve the overall quality of care. The extensive growth in the available health data, accompanied by powerful computational resources, has the potential to provide profound new insights into healthcare. Investment in the interpretation, rationalization, and management of health data has the far-reaching potential to revolutionize personalized treatment, the overall quality of care, healthcare efficiency, or the development of novel cures.

With the progression of health data statistics and computational methodologies, it is expected that the pace of healthcare innovation can be accelerated by utilizing the routine health datasets. This reflects the potential of next-generation health statistics and data-driven methods and the far-reaching impacts on advanced healthcare in the twenty-first century. An increase in the uncertain climate for medical services suggests the need for more innovative methods of health outcome prediction, which can be a valuable tool for improving the efficiency of doctor visit scheduling, or anticipatory delays in managing bed allocation on wards or intensive peaking in stocks of essential medical. Broadly speaking, efforts include treating statistics and predictive analytics as a means of living prudently with uncertainty in the health community [1].

1.2. Fundamentals of Predictive Analytics

Healthcare data have exploded in recent years, with about 30% of the world's data in 2011 created by the healthcare industry, and the amount of health data doubling every eight years. Such large data volumes represent opportunities for healthcare providers to make better decisions, but the expertise to understand these data and make meaning of it has not expanded at the same rate. This in part due to a limited understanding of predictive analytics and lack of resources to implement new technologies. Nevertheless, hospitals can no longer afford to ignore the opportunities presented by their data; rather they must harness their data to achieve sustainable operations and healthcare services.

1.2. Fundamentals of Predictive Analytics

The purpose of a predictive model is to forecast the behavior of a response variable based on the values of a set of predictors. In healthcare, the response could be the risk of a condition, disease, or event, and the predictors could be demographical data, vitals, or previous lab test measurements. A range of techniques exist for modeling and judgments can be made on which of those techniques, 1) Don't require extensive knowledge in machine learning or statistics, and 2) Are amenable to a wide variety of healthcare data. Typical methods that can be developed, or are currently readily available, include logistic regression, naïve Bayes, and random forests . However, it should be noted that with complex health problems requiring sophisticated models, it is wise to collaborate with data scientists.

2. Applications of Predictive Analytics in Healthcare

Among the most rapidly developing industries in the recent past, healthcare is unique in many aspects. Being one of the world's largest and oldest industries, healthcare sustains an immense number of lives. The total expenditure on healthcare in the US has reached approximately 3 trillion dollars, comprising 17.5 percent of the country's GDP. Fast growth rates have also been observed in the healthcare sectors of Europe, Asia (especially China), and Australia. Taking Australia as an example, the total healthcare expenditure has reached more than 121 billion dollars [2]. There have also been many gigantic dole-outs from the government. There have been an increasing amount of people involved in healthcare, including health economists, clinicians, health policy makers, technicians, and information systems engineers. There are also undergraduate medical students and interns to enforce one of the oldest Hippocratic oaths.

Healthcare is one of the largest service industries in the world, involving many people either as employees in healthcare systems or as consumers of healthcare services. Approximately 16% of the US employment establishments are in healthcare, employing more than 6 million people. One aspect of the healthcare industry that has always interested researchers is the effective delivery of promoted healthcare services when needed. It also comprises the improvement of healthcare service delivery, including the efficiency of the services, their availability, and their quality. With the availability of datasets that contain a large amount of personal health information, further studies were conducted on the effectiveness of implementing personalized wellness programs. Often involving complex systems with both uncertain parameters and exogenous variables, a general description health care state estimation involves the following premises, tasks and approach. Tasks include estimating the current and future states of health care variables based on measurements of the system behavior. This generally assumes that the underlying physical phenomena are well understood, and it involves estimating variables such as model inputs or parameters, which are not directly observed [3].

2.1. Disease Diagnosis and Prognosis

Disease prediction is of central importance to living a healthy lifestyle in the modern age of an information overload. It is the hope that by turning raw health data into patient insights and making predictive disease modeling more accessible, individuals will consume health resources more carefully. A web app with implemented predictive disease modeling tools is released to accomplish this objective. This web app features a core engine for predicting a patient's risk of contracting certain illnesses in the future. This engine is accompanied by the patient’s counterpart, which generates insights about the modeling of diseases in narrative form. A large-scale randomized experiment is conducted to ratio the disease prediction stories to a variety of synthetic stories on patient health behaviors. This experiment aims to assist individuals in establishing health-related behaviors that considering their future disease predictions, hence minimizing their chance of getting sick. Moreover, data from trial sciences are analyzed to help behavior nudge insights.

A trend like the story today, could be an easy start. By the time the dinner is finished, the tummy's ache becomes critical. There's no good smell for years. The nose's tricking the brain is still valid, though. But came the realization like yesterday, the body argues. All the dear ol’ friends from back child time are abandoned. None of them are invited all the three times in a day. Just because of the eyes' beholding, the mouth achieves the greatest integration and the tummy seems to comply. But the profit achieved seemed too hilarious to be true. The present's forever decisive to all of life's precious journeys. It proves to have kicked off the treasure to the wrong cavern. Engulfed just small portions, the belly groans. But even for such a few, giving it like always ends up making one forced to think back then when for a dumb head it triggers a laughter [4]. It was commonly said: “May the day never come when the joy of hearing those words read aloud is lost.”

2.2. Treatment Optimization

Optimization of individualized treatments poses a significant challenge because it requires simultaneously selecting an appropriate treatment from a pool of options and for an individual patient. A booming healthcare challenge is that existing capacities in the healthcare system are often overwhelmed by an increasing demand for medical services, which is induced by an aging population and an increased incidence of chronic diseases. This leads to longer waiting times for patients who need treatments, many of which are a progression of illnesses, such as various cancers. Worse, conventional “one-week and see” treatment scenarios lead to unnecessary treatments and cannot guarantee best treatment choices for individual patients [5].

With the increasing availability and improvements in electronic health records (EHR), learning-based methods that learn from the large patient databases are increasingly feasible. The primary goal of the learning task is to analyze historical EHR data to build a prediction model for predicting outcomes or estimating risks, and, given current knowledge, the model can predict how patients' outcomes will change after the receipt of alternative therapy treatments. Importantly, the increasing volume and variety of EHR data can lead to the development of prediction models with enough capacity to capture the underlying complexity of treatment response. On the other hand, advances in metrics computing and computational algorithms enable personalization or the learning of optimal treatments for patient groups, hence providing a means to leverage predictive models of treatment outcomes [6].

2.3. Resource Allocation and Management

Disparities in health systems rooted in social and economic structures have proven difficult to eradicate. Low- and middle-income countries (LMICs) fail to receive essential public health goods on account of lacking necessary physical, financial, and medical resources. In this text, a framework is presented in leveraging country-level health data in predicting future patient outcomes based on past treatments during hospital admission. By using patient health records (PHR), health data such as past medical history, surgeries performed, medicine prescriptions and the discharged diagnoses of the patient are used to predict future health outcomes. The feasibility of the proposed framework is demonstrated on PHR data from American state hospitals. This modeling method can be employed by health organizations to forecast further patient treatment and provide opportunistic or preventive care to patients most at risk. Proposed predictions are particularly beneficial when such care has limited resources, i.e., in poverty-stricken environments much like LMICs [7].

3. Challenges and Opportunities in Utilizing Health Statistics for Predictive Analytics

In this data-driven era, the world is witnessing a deluge of data that is being churned out at a rapid pace and in huge volumes across various domains, with healthcare being one of the most prominent areas among them. The data in healthcare systems is in the form of electronic health records (EHRs), laboratory tests, medical imaging information, and genetic data. This large-scale data is referred to as health data or health statistics. Predictive health analytics is the discovery and communication of meaningful patterns in health data, and the use of those patterns to foretell, or predict, health and sickness. It plays a vital role in transforming healthcare outcomes and bringing forth personalized treatment plans. Health statistics still offer great opportunities in the informatics research areas even though challenges exist. It is a difficult and challenging research topic to utilize extensive health statistics for the meaningful and effective predictive health analytics. A number of issues need to be addressed in the upcoming research in order to promote the applications of health statistics to personalized healthcare. Despite the tremendous potential that big data holds, there are a number of challenges associated with it. There is an understandable fear about the sharing and storage of massive amounts of health related data, in part because of the risks of revealing the identity of the patient based on various sources [8]. However, despite the challenges that they need to overcome, the advanced analytics that are promised through big data offer tremendous opportunities for most stakeholders in the health care industry--patient, provider, and payer--ranging from improved personalized treatment options for the patient population, and health outcomes for individuals, to refined effectiveness of treatment plans, and operational strategies for managing patient populations. [9][10][11]

3.1. Data Privacy and Security Concerns

With the development of the Internet of Things, potential health indicators produced by sensors are connected to the cloud that collect data continuously to detect symptoms of patients' sub-health conditions to avoid possible chronic diseases. Based on this data, new applications that can foresee patients' chronic diseases or other health risks are being developed. Patients could subscribe to a cycling service in case of a diagnosis of certain health risks. Nevertheless, data generated by models outside the health center cannot be used for triage processes on the corresponding pathology. Therefore, there is a need for the emerged symptoms from such patients to be detected in their health center’s system. To implement this, cloud-based notification mechanisms would be very important. Definitions of symptoms can be defined within the cloud so that they could be constantly updated given the vast cases in a global perspective. This is seen as important support for health systems, but there are patient privacy and sensitivity issues. How can this novel approach that could be adopted from preventive patient perspective and in favor of personalized healthcare be improved? As a result, what novel challenges and open issues does it raise in healthcare technology that encompasses IoT and cloud infrastructures? The following outcomes were intended to be obtained: (1) emerging health indicators and why early symptom detection is crucial for the future are briefly introduced; (2) a new vision is proposed based on a series of automatically-generated procedures and services driven by cloud-based health systems and data, and these are discussed; (3) as actual implementations can take a long time, early needs and a time-saving fridge for healthcare systems and policymakers are conveyed to prepare infrastructure and data policies given the future assumptions and discussion [12]. Decreasing the utility of data is due to the privacy-preserving analytics algorithms and may affect modeling predictive pathways and optimization procedures. Advices should be enacted under realistic circumstances. Prescriptive practices, however, are at an early-stage of concerns. Recently, there have been increasing efforts in healthcare pertaining to the online e-health marketplace. Several companies have developed new business models on e-health suites providing online medical advices and teleconsultation services for patients. Understanding such patient behavior is significant for strategizing pricing, consultancy, and e-commerce. However, there is a potential for great data abuse, resulting in a conflict of interest with patients or with a community. The sharing of complex data also generates ethical and legal problems for data holders. Navigating such ethical dilemmas is at an early point, and change can occur exponentially as biomedical data expands. Certainly, further discussion, including legal analysis, is needed among legislators, data providers, and policy makers. Nonetheless, the potential cost of 'quick' data examination is highlighted. The exploitation may occur for political reasons or to put pressure on certain actors for the benefit of the industry and not the community. Efforts to resolve these dilemmas will put pressure on democracy and transparency. Therefore, e-catalogs should be handled with respect and accuracy, including proper usages of methods and tools for data inspection. Broadening the consultation base and involving more actors such as civil society and consumer protection organizations is also advised. Evaluation of forecasts must be conducted in a neutral manner and additional insight should be developed. In particular, with regard to individual risk forecasts, such predictions should be combined with guidelines on modifiable risk factors and drugs in order to optimize treatment course [13].

3.2. Data Quality and Integration Challenges

Health data within healthcare systems is highly distributed, but available statistics are commonly aggregated and derived well after the data is collected. The promise of predictive analytics in healthcare is to empower healthcare directions towards more responsive patient outcomes [14]. Source data needs to be aggregated, transformed, and cleansed prior to conducting analyses. Health data presents challenges in this area as personal privacy and ethical concerns restrict data sharing, and data quality may be influenced by a variety of external factors such as health insurance policy changes. Difficulties in procurement of data, and infrastructural limitations for collected data are addressed, namely the implementation of these learning algorithms at a community health center that serves a population which is both underprivileged and uninsured.

Reflexion Health provides Physical Therapy services with their own health professionals and session equipment to patients in their own homes for greater convenience and adherence. All physical therapy sessions in the Reflexion Health database are analyzed, evaluating the acceptance rate of patient-generated data, and stratifying patient characteristics which correlate with good adherence, valuable information for designing subsequent healthcare services. Furthermore, a collaborative data sharing policy is outlined, as well as an organizational data infrastructure that allows for the collection and analysis of granular patient health data from the frequency and severity of hand tremors to data.

3.3. Interpretation and Implementation Hurdles

For most healthcare professionals, predictive models are just a “black box” [4]. Doctors and bedside nurses might have hard time grasping what the output of a computerised prediction really means, e.g. depending on the severity of some vital signs and previous medical history, a future surgery might be listed as “high risk” simply because severe states are necessary, not because an unwanted outcome is more likely. Surgeons and anaesthetists might discard the model-assessed risk foreseeing a “high reward” case, instead just as their colleagues in other fields, they can be overly confident in their ability to quantify risk intuitively [15]. Unfortunately, statistical models are generally more reliable than experts’ opinion, accounting for every piece of data and maximising the prediction accuracy. After all, statistics is the science that idealistically aims to tell what is the occurrence likelihood a future event given a combination of states of input data. The transition from the abstract world of numbers and figures to the real and far more complex reality of wards and patients is often thick with pitfalls. Oddly enough, the required translation becomes less intuitive as predictive models get more accurate and computationally demanding. From a theoretical standpoint, a large set of complex rules is needed, precluding the recognition of simple patterns among vast sets of variables. A common approach for the interpretation of a black box system is to feed it synthetically generated statistics, so the effect of every single feature can be calculated. Despite being very insightful, these do not apply in the real-life world.

4. Machine Learning and Artificial Intelligence in Healthcare

Innovative advancements in machine learning (ML) technologies and artificial intelligence (AI) analytics algorithms, including predictive analytics (PA), are generating new opportunities for healthcare providers by allowing efficient examination, interpretation, and reasoning of large amounts of health-related data for recognition of patient risks, forecast of results, improvement of therapy, and reduced pejorative results [16]. This narrative review provides a synopsis of PA predictive models’ functionalities, applications, benefits, limitations, and challenges, offering a basis for enhanced understanding and interdisciplinary cooperation between physicians, nurses, IT professionals, policymakers, scientists, and PA analytics sectors in fostering healthcare transformation. PA is as a subset of AI that focuses on the processing of historical health-related data to generate numerical insights, such as future trends, occurrence risks, and optimal decision choices, directed towards the accomplishment of predefined business goals. Four types of PA models have been created and confirmed as pertinent by health scholars and data scientists: disease risk; therapy assistance, effects, and adjustments; forecast of results, adverse events, and re-admissions; individual illness analysis and optimal treatment personalization (e.g., stratified, casual inference). These models have been applied in numerous studies to diverse medical conditions for distinct medical examinations and will be demonstrated as such in this review. A predictive, accurate, and natural forecast technology, particularly one that transcends complexity (i.e., AI/ML-based PA analytics), can anticipate adversaries and permit proactive choices against them. However, the technical complexity, model inscrutability, and data sensitivity of these advanced systems have fostered a lack of physician buy-in, nuanced regulatory discussions, and more subtle ethical considerations. Moreover, these studies seldom echo on the theoretical limitations, case-specific behaviors, and externality dependencies, and thereby extrapolation to a broader understanding of penetration is made complicated. [17][18][19]

4.1. Types of Machine Learning Algorithms

Health statistics and predictive analytics are an inevitable pair that will transform healthcare outcomes worldwide. Paralleled with the advent of supremely digital healthcare infrastructure, substantial health data has been collected in monitoring patients, diagnosis, and treatments [20].

The substantial data generated, however, creates unparalleled challenges in data management, analytics, and decision support. In this context, together with the case study of predicting hospital readmission rates as an example, machine learning algorithms and concepts for health predictive analytics are briefly introduced and demonstrated, which may potentially transform the healthcare outcomes of patients and hospitals.

In healthcare, the benchmark of predictive performance by prospective analytics is the potential of early treatment and interventions before the clinical deterioration of the patient [21]. It requires a thorough analysis of health data and a reliable model to uncover underlying patterns hidden in the thickness of the dataset, considering the potential of intricate interactions and multivariable data assessed in clinical practice. Machine learning techniques, such as ensemble learning algorithms, deep learning algorithms, and gradient boosting algorithms, are considered as a state-of-the-art technique to learn from complex data allowing the analysis of a broader range of features and relationships. Devices for the volume and variety of health data from different modalities are engulfed. While focusing on health data analytics, the data preparation and analytics steps are illustrated. In the discussed case study, the health data includes structured data (i.e., patient demographics, diagnoses, and time series of vital signs measurement) and unstructured data (i.e., textual data). Machine learning algorithms are illustrated on how to leverage diverse health datasets for analytic purposes. [22][23][24]

4.2. Deep Learning Applications

Artificial intelligence (AI), particularly deep learning (DL), has shown significant advancements in predictive analytics leading to multiple efforts in healthcare applications. DL is used in applications to analyse complex data and has improved the performance of voice and speech recognition, visual object recognition, game playing, and natural language processing. The AI scope includes prediction, patterns and anomalies search in data, detection and classification of diseases, and recognition of signals of a normal behaviour and drug-target interactions. Time to event, a common form of clinical data representing survivability, can also be processed to extract valuable prognostic indicators. For medical evaluation and monitor deep image analysis has improved task-specific algorithms published for image-based predictions. Negative effects of medical treatment or interventions methods can be fixed with pattern detection through learned deep training.

Accurate diagnosis is a milestone for the personification of medical treatment. The potential of improving prediction outcomes of complex diseases with various deep learning architectures and advanced time-based features is illustrated. Breast cancer metastases detection studies on the testing data from the CAMELYON dataset show that seven deep learning algorithms trained for the detection of metastases in lymph nodes of women with breast cancer outperform a panel of eleven pathologists. Cardiac disease diagnosis is figured out on the test dataset from the ACDC challenge that includes echocardiography videos of individuals with and without cardiac disease, altered in both training set and normal images. An echocardiography system with a novel temporal-stretching convolution layer is presented for the automated diagnosis of the cardiac disease. With echocardiographic view classification accuracy of 80% AUC, complex temporal patterns are learned directly from the tensor representation of echocardiography videos. Early prediction of Alzheimer's disease was achieved with the tester dataset from the OASIS 3 database. A deep learning system is described that demonstrates improved early prediction of Alzheimer's disease utilizing brain PET, a subtle signal of neurodeterioration, providing enhanced opportunities for early therapeutic interventions [25].

5. Ethical Considerations in Predictive Analytics

The use of healthcare data from different sources can provide a detailed picture of an individual's health status, and ultimately what their future health status could look like.

Predictive analytics can be used to leverage health data effectively, improve clinical decision-making, and inform health policy, so European Union Member States should develop a clear evidence-based strategy to ensure actionable information is derived from health statistics. Backing this should be high-quality, potentially the real world data source for analysis and shared with policymakers, researchers, healthcare professionals and the general public to enable informed decisions. A multitude of predictive analytics tools are developed to explore the associated potentials and challenges, including significant differences between regions in 3-year predicted hospital admission rates for 65-74 year olds.

Furthermore, there are other possibilities arising from the increased use of health statistics in health related applications, such as the prospect of developing personal health record-based predictive models, enabled by the growing interest in personal health records. Personal health records allow patients to view, download and transmit their medical information and meaningful use metrics show an increasing number of healthcare providers using personal health records or particular functionalities of personal health records, e.g. in patient appointment reminders or delivery of diagnostic test results. The collaborative nature of this approach could support the development of a new way to position health statistics within European Union health policy, expanding their potential use, particularly on the international level. Therefore, this exercise could be regarded as a rightholder's view of a high value, but underused, resource. Since health statistics play different roles in different health systems, a broad overview of the informational functions of different health statistics systems is warranted for context [26]. Using several member states as examples, the landscape of health statistics in Europe is described alongside an overview of the current international environment that surrounds health statistics.

5.1. Bias and Fairness

The harvested health statistics are harnessed and projected into the future with the utilization of a predictive algorithm. Hence predictable patterns may be of great guidance to the health professionals in their efforts to enhance healthcare outcomes. Results in similar domains, where individuals were encoded in search engine queries or credit application, unveiled that it is attainable to anticipate not only health condition but also observation-related demographics of these individuals [27].

As disclosed in the media, there are substantial apprehensions that the very same attempt to enhance some individuals could lead to unfair preference towards other individuals. The concern is raised that wide-spreading predictive analytics in health provision could pave the way for the discrimination regarding some individual groups within the culture. The lodging of these issues is not plain in the practice as they usually come veiled in semantic intricacy, exacerbated by jointly philosophical and technical aspects. Inspired by the extensive works, numerous are only incognisant of some elements of the process. There are usually browsed as one of two fractional issues. First is the problem of justice, equitable apply of the law, concerning merit and equity. Second is simultaneously construed as this and as theoretical fairness postulate and it has been protracted what is a fair tense as learnt by machine. Technology evolves and fair algorithms are intended not to evaporate inequities. Beside these recurrent concerns, a comprehensive piece states uneasiness particular to the predictive apparatus, but greatly delimits these concerns are infeasible for the distinctive implementer. It is anticipated in the very next years there will be veritably wide-spreading public apprehension as well as emergent works on this subject. Effort has been made to unclose and shift understanding towards additional distinguishable, actionable fair apprehensions and possibilities.

5.2. Transparency and Accountability

Developing a platform for predictive analytics is only the first step in propagating data-driven practices for making better and more informed decisions in health care. Fair analytics means that any user can access and process the data, and understand the resulting analysis without any proprietary modeling code. Generally, the data used to create a model are withheld. The choice of the data used can be motivated by the need to provide a transparent means for verification and validation of results. A model that is fair in that sense is maintainable because all the analyses leading to its creation are open to scrutiny. In the context of predictive analytics, transparency is interpreted in the sense that any interested third party can use the model as long as they have the tools to run it. With predictive modeling, the meaning of transparency is that the model output includes not only the prediction, but also how the prediction was obtained (what features were used, and their contributions).

Such data are consisting of de-identified patient outcomes, costs and ratings (e.g., how patients rate their hospital stay, how hospitals are rated by state inspectors). This presents a unique opportunity to analyze the patterns and trends in the health of one of the largest states for a period of one decade. Common ailments can be quantified, such as the number of patients admitted with the common cold, flu or heart attack. Time evolution can be studied for seasonal diseases like the flu, or how hospital costs evolved for the treatment of the flu. Data on expenditures by hospitals are available in a voluntary and mandatory form. In addition, hospitals have their own earnings, so there are many figures about money coming out and going into hospitals. So making sense of the flow of funds can be challenging – it is difficult to sort the fiction based on separate data releases. However, there is no public data of such origin. On the contrary, much information about medical costs and various procedures is publicly available through very large bills one might have received from a prolonged hospital stay.

6. The Future of Personalized Medicine

An end to the one-size-fits-all healthcare?

Personalized medicine — the idea that individualized treatment can be tailored to a specific patient on the basis of their genome, phenotype, or other omic data — broadens the scope of disease detection and prevention and offers significant benefits to the quality of clinical practice by enabling timely, effective, and evidence-based care. It also addresses economic reasons. Early detection can prevent disease. Preventive approaches are cheaper than late-stage interventions. Accurate risk assessment can enable selective prophylactic schemes. Additionally, it is anticipated that the bioinformatics approach will improve the overall efficiency of medical services [28].

With the rapid advances of high-throughput biotechnology, biomedicine is broadening its focus to bridge the gap between genetics and disease. From big data analysis in whole genome to omics-based techniques, personalized medicine is evolving on an unprecedented scale. Procedures, such as electronic medical records and national DNA biobanks having unique DNA-phenotype associations, contribute to the advancement of personalized medicine. Nevertheless, numerous challenges lay ahead. The question of how to interpret the uncovered big data while avoiding overfitting is a central issue. Solutions include improvement of biomedical knowledge, construction of bi-layer models, and incorporation of pathway analyses of omics data.

6.1. Genomic Medicine

One of the cornerstones of the upcoming forecast health reformation lies in the deployment of new complex health statistical methods. Predictive analytics, in particular, obtains knowledge patterns in clinical and health data so that inform decisions can be determined in agreement with them [29]. In the collaboration of the National Health Services with universities and private corporations, these methods can play a crucial role to rethink future healthcare and empower novel personalized therapeutic systems. The statistical–mathematical methodology can be designed for those applications. Specifically, the most relevant domains are genomic medicine, prenatal health, drug use optimization, and stationary assist on medical and care teaching. Some open net issues at the intersection of medicine and predict technologies are also approachedimiters.

6.2. Precision Oncology

Each day brings forth new statistics about different diseases. Therapies seem to outnumber the so-afflicted. Even though these statistics are dense, good practices are nontrivial. Precision targeting in medicine can uniquely handle each specific case. There is a unique challenge in making sense of this data. While the field aims to make statistical meaningful, it doesn't always succeed. A bioinformatic approach involving protein interaction network alignment can use statistics to prioritize the most promising treatment targets. During the course of precision targeting, a lot of questions will come up. Some are simple idiosyncrasies, such as why two antonyms, thermally well-established and statistically significant, share the common synonym substantial. Extra efforts are often put into cleaning and refining the statistics. In the above case, this involves reobtaining and reanalysing how cancer driver genes tend to be more network central than their synthetically lethal counterparts. This revitalization, in turn, yields a counterexample where complete network information is unnecessary. Digging deeper, a more subtle question presents itself. Pondering whether network centrality in multi-interactomes aligns with network centrality in identically sampled single networks motivates introducing stimulated filters. Besides identifying a rigid mechanism to turn statistics into biology, these filters capture the centrality as a function of the neighbourhood.

Like many examples of language, its word pairs form a spectrum where both strong and weak ties coexist. While part 5 may seem overly pedantic, it's included with inadvertent inspiration and inevitable stimulations in mind. Just as a few lines of translation encouraged Newtonian tranquility, it's hopeful that better filtration can impel a more pervasive resolution to this precise question. Moreover, upon movies, there's quite a slew of tumors, a mosaic of patient types, flocks of moieties. There's quite a slew of pharmacology options. With personal genomes sequenced, it's hoped that molecularly-informed decisions from the vast repertoire of omics data will guide treatment choices. This turn to precision medicine is met head on in the realm of cancer. It's hailed by some as a paradigm shift, it's employed in the broad sense connoting treatments tailored to the specific genetic make-up of an individual's cancer. In clinical practice, this means sequencing the DNA from the patient's tumor, spotting the culprits, the "driving" mutations, and inspecting an onboard databank to unveil existing medicines that can thwart such mortiferous alterations. Better still, it involves discovering drugs in the works or compounds that can be re-purposed. The promise of precision oncology is obstreperous. Despite a universe loathing change, almost have been buoyed into biotechnological enterprises, supporting masterpieces and private-sector Dodos alike. Beginners and veterans alike have conceived breathtaking projects, platforming a new generation of wizardry appliances. Importantly, agitated by the preach of progress, teams have melded into coalitions, groundbreaking inquisition hubs have come to life, aspiring tongues profess to the uni-search of knowledge, and perhaps cures. Observing from this pandemic flood of happenings, a spectator may so naturally swim from marvel to rapture, but then, between the crests there lurks repose and reflection. Two questions arise then this feature. Just how alien is the overdose, and what of the doctrine unseen by cover? [30][31]

7. Case Studies and Real-World Applications

Health statistics are a significant resource of evidence in exploring healthcare outcomes and personalized medicine. Using predictive analytics, findings can lead to better diagnoses, management of diseases, lowering readmission risks, and prescribing optimal treatments. First, a learning-based evaluation strategy is proposed in the feature selection process, and then, a novel statistical sampling (Poisson-difference) is presented to generate synthetic datasets based on statistical variation for evaluating stability of the feature selection. The effectiveness and stability of the proposed feature selection are demonstrated through comparing the prediction performance of four well-known classifiers on 14 healthcare datasets.

There are 14 studies revealing various aspects, such as the evolution of disease status, the influence of lab tests, and the impact of missed treatments, in predictive analytics of healthcare. Particularly, most case studies focus on the utilization of inpatient records, which primarily describe patient management at the acute time (or in an emergency setting), and only binary or count diabetes, liver diseases, and heart diseases are studied [32]. However, the predictive analysis of various aspects in the healthcare field is still inadequately explored. This study intends to reflect on existing analytics and assess what can be done to further investigate the ability to utilize health statistics that predicts essential elements surrounding a wide range of healthcare outcomes.

7.1. Predictive Analytics in Chronic Disease Management

Chronic diseases and poor lifestyle choices impose significant threats to the health and well-being of a national population . Increasing numbers of patients with chronic infections result in higher healthcare costs and increased use of medical resources. However, many people with chronic infections lack adequate medical attention and assistance to avoid the deterioration of the disease. Therefore, it is important and also difficult to explore an efficient way to manage disease for patients with chronic infections. Fortunately, the widespread use of Internet technologies and wireless technologies have brought new opportunities for chronic disease management.

A new way of chronic diseases management system is proposed leveraging the combination of statistical approach and mobile pervasive technologies. A multidimensional statistical metric is present that simultaneously assesses the degree of stability of health related parameters and the correspondence between this temporal regularity. The metric is able to predict the vulnerability of a subject to chronic disease. This statistical metric is used as the indigenous function of a mobile chronic diseases monitoring system. The system gathers periodic healthcare data from wireless and non-invasive devices: blood pressure, blood glucose, heart rate, temperature, weight, and other human health constants. The healthcare data are statistically analyzed and filtered by such predictive model. In opposite, feedbacks from the system are given timely to alert, suggest food, prevention, and encourage physical activity.

7.2. Personalized Treatment Plans

The advent of routine collection of Electronic Health Records (EHRs) has produced a wealth of health-related data that can be exploited for predictive problems. As a result, there has been a significant interest in predicting the effects of alternative therapies when planning to treat a patient. The goal of this work is to learn decision rules from historical data that, given an individual's healthcare trajectory, will form the basis of individualized treatment recommendations. Unlike most works in the literature, which deal with the development of stationary treatment rules and assume that the data follows a randomized trial design, this work is interested in learning personalized treatment rules from time-varying covariates via observational data. The treatment plans are dynamic, and patient's characteristics in terms of the variables observed at each time are accounted for. Developing treatment rules under guidance from observational data is particularly challenging due to possible confounding effects. A method is proposed that tackles this problem by an innovative approach based on non parametric modeling and kernel based regression, taking advantage of the availability of large quantities of data.

Randomized clinical trials (RCTs) are the study design of choice in the medical sciences, for drawing valid inferences about a potential causal relationship between a treatment regimen and patient outcomes. Strictly controlling the assignment of the treatment using the randomization mechanism has made it possible to estimate the average effect of the treatment on the treated subjects. However, the advent of personalized medicine, also referred to as precision medicine, has questioned the value of the average treatment effects as estimated by RCTs. The major concerns are that, given the heterogeneity within the tested population, there is the risk that some effects may not emerge, or that the estimates made are too imprecise for practical use. There are several examples in medical practice, in conventional fields, in which RCTs perform poorly as a system for establishing evidence. Randomized studies have also been conceived to be less effective in the context of non-therapeutic data, like diagnosis [33].

8. Data Sources and Collection Methods

Participant data can be collected from a wide range of sources, both manually and electronically. A list of different data sources and collection methods is provided in the following paragraphs. During exercise stress testing, patients often have a 12-lead electrocardiogram recording taken, and sometimes it is done during recovery after exercise. Such ECG recordings can be used to estimate the volumes of ventricles of the heart. Researches have worked on extracting these ETTs using two approaches, either from continuous ECG recordings in contiguous segments over time, or from three ECGs recorded during different points of the ETT. Resulting ETTs are used to predict the occurrence of left ventricular dilatation within one year of the ETT examination. Similarly, electronic health records containing demographic and lifestyle data also have the potential to provide deep insights into disease mechanisms when analyzed by artificial intelligence algorithms. Diet, which is known to play a crucial role in determining risk for certain diseases, can be quantified and translated into features indicative of underlying dietary habits. Such converted features capture several aspects, such as the co-occurrence of healthy and unhealthy food items, the variety and diversity of the diet, and various measures of the unhealthiness and the timing of food consumption. The data collected by simple questionnaires could also play a central role in this context. Binary matrices can be constructed, where each row represents a participant, and each column indicates participation in a particular activity. Deep neural nets can then be implemented on that data alongside AutoEncoders, which enable the detection of clustered participation in different activities.

8.1. Electronic Health Records (EHRs)

The data elements eligible for consideration in any EHR-based epidemiologic study are those documented during healthcare service provision. They can be grouped into: demographics, such as gender and age at date of service and several categories of location indicators; vital signs, such as height, weight, temperature, blood pressure, respirations, oxygen saturation, and heart rate; diagnoses and disorders; procedures, therapies, and services performed on or provided to the patient including physical and occupational therapy; medications, therapies administrated concurrently, or requested to be taken concurrently; laboratory orders and tests; vaccinations and immunizations. Described quantitative and qualitative attributes of EHR data rely on measurement processes within the clinical enterprise and on documentation tactics of individual healthcare providers. In a vocabulary-depth analogy documentation activity can be understood as the sum over time of internal and external documentation tactics of an entity. Documentation of healthcare encounters is essential to quality of patient care and insurance reimbursement for services provided. It is empirically estimated that, on average, American physicians spend 1–2 hr of a common 8-hr workday documenting clinical encounters. This time burden is not lost on healthcare providers. Many view EHRs as 'engines of oppression', and scoring low on job satisfaction surveys has been associated with high volume of documentation duties. Automated documentation technology such as speech recognition dictation software, scribes, and templates can help with electronic documentation efficiency. Reflexivity of service provides to patients may also improve workflow productivity [34].

8.2. Medical Imaging Data

Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension

The rapid growth in the use of medical imaging for clinical practice has led to the accumulation of medical imaging data. This data can be utilized in imaging-based observational research. To do so, a systematic methodology is required to manage, harmonize, and standardize medical imaging data. This study aims to develop a medical imaging extension of the Observational

Medical Outcomes Partnership (OMOP) common data model (CDM) and characterize and validate mapping rules for multiple imaging data sources to the OMOP-CDM [35]. Development of the radiology common data model for international standardization of medical imaging data. Medical imaging is widely used in clinical practice. The imaging data could provide essential information for patient care and is critical for sound care delivery. Health systems and hospitals around the world are generating large amounts of imaging data and utilizing the data for clinical and research purposes. However, the accumulation of large data volumes and the complexity of addressing diverse imaging modalities and their findings are barriers to using medical imaging data sources for secondary purposes.

This study demonstrates a repository of multiple-years' worth of clinically acquired DICOM imaging data extracted from an academic hospital picture archiving and communication system. The Washington University School of Medicine echo of MINA-T, a pictology repository containing 1.3 billion images encompassing 6.9 million radiology and 21.7 million non-radiology image series, primarily in DICOM format and de-identified in compliance with the Health Insurance Portability and Accountability Act. RSNA clinical trial processor is an informatic tool that facilitates the de-identification of DICOM imaging data. Conceptual work was performed to identify a set of data structures representing hospital pictology studies, with specific attention to radiology reading sessions.

9. Predictive Modeling Techniques

Predictive modeling techniques are increasingly used in research and clinical practice [36]. It is known that, using health data, some conditions could be predicted accurately within a certain time period. However, as the generated data grows larger every day, health data becomes more and more important every day, both for research and practice. The biggest benefits from health data can be achieved by transforming it into meaningful information using statistics, indicators, and mathematical modeling. Thus, this review aimed to show the transformation of health data into meaningful health results.

A clinical prediction model is basically used to predict a specific health condition or disease for each individual. There are two types of clinical prediction models: diagnostic and prognostic prediction models. A diagnostic prediction model estimates the probability of an individual currently having a specific health condition. Conversely, a prognostic prediction model estimates an individual's probability of developing a specific health outcome over a specific time period. High-quality clinical prediction models have the potential to be highly instrumental for clinical practice, health services planning and evaluation, and for health policy areas such as the allocation of resources, incentives, and performance management. On an individual level, clinical prediction models are useful to identify people at high risk for a certain outcome, who may benefit the most from (preventive) intervention, and to efficiently and hence cost-effectively empower people with an informed choice of actions on how to reduce their risk. Clinical prediction models may be developed using exclusively or mainly routinely collected clinical data, but might also integrate other real-world data. Model building involves the selection, coding, and processing of all candidate predictors, the selection of its functional form, as well as the type and degree of stratification of all candidate predictors, and the evaluation of the modeling process, validation performance, and clinical usefulness. Ways to undertake these steps, as well as major issues and recommendations to increase the validity, quality, transparency, and reporting of clinical prediction modeling studies are discussed.

9.1. Regression Analysis

Even when the data generating mechanism is known, the properties of the training data, models, experimental setup, and candidate variables can change between inference and prediction settings [37]. This study compared predictive modeling against standard explanatory modeling using random forests for the task of recovering generated signals in cross-validated common real-world medical datasets. There were five reasons why interpretation and out-of-sample generalization diverged in studying both simulated and empirical biomedical data.

Random Forest (RF) regression was used to summarize performance variation in ordinary least-squares regression (OLS) and Lasso as a function of the experimental properties of the datasets. This allowed the researchers to compare OLS and Lasso against the same reference model in all cases and to identify which experimental factors across 43 datasets impacted differences in recovery performance. The number of truly relevant variables, which significantly impacted recovery performance in both OLS and Lasso, was also varied across datasets. Model violations, which were defined as explanatory variates being squared to generate the dependent variable, were important for differences between analyses using OLS and Lasso but were never fully captured by RF regressions.

9.2. Decision Trees

Industry research has been an early adopter of data mining. Decision trees developed in the area of decision theory were introduced at the same time. A decision tree (DT) is a representation of a multivariate function that can be understood and used in everyday life. Since then, despite the opinions of its opponents, the DTs have progressed to the point of becoming a practical and easy-to-use tool.

The DTs that the general public mostly know are based on the CART (Classification and Regression Trees) algorithm. CART is the acronym for an easy to implement method of induction used to build decision trees recursively. DTs have been developed independently of the ID3 and CART algorithms, and different algorithms have led to the creation of trees with different characteristics. The ID3 algorithm creates trees based on the entropy of the information. The topic of these studies is not the decision tree as a general concept, but the development and use of those methods that are capable of correcting the flaws in the decision tree used in the classical decision theory. That model, which is currently widely used in medicine, has several shortcomings. Some of these shortcomings are the lack of support for continuous variables, the tendency to produce overly complex models, the inability to develop rules for the model, or when there are more than several possible classes. Another one of a number of patients treated at a medical institution may also mean that there are missing values of some patient data. Medical decision-takers may fall into several groups. Broadly speaking, possible users of a decision tree are physicians making treatment decisions, public health authorities, but also patients, if they are oriented towards personalized medicine. On the other hand, a doctor of any medical institution with a large number of patients in need of care will want to have a prediction tree covering all of them and will want to use it in practice. It can be expected, moreover, that responding to currently fashionable concerns for patient-oriented medicine will not be left out of the benefits of IT advances in healthcare. Also, if it were possible to analyze several different trees and some would manage to develop a clear statistical pattern, the predictive efficiency of such implemented trees would certainly increase as well. This important research developed a framework for the use of DTs in medicine and biomedicine. In discussing medical uses of the DT they describe several interesting possibilities in this context. On the basis of these works, in this part, some of the most important possibilities of licensed use of the DTs model in medicine have been discussed. The aim was to test, in the set of diseases treated at the Upper Silesian Center for Heart Diseases in Katowice, the prediction and classification ability of the Bayesian network, feedforward neural network, k-nearest neighbour algorithm, logistic regression, and decision tree models, and to find the best method for this inductor. Plasma lipid concentrations are a common measurement used to assess the risk of cardiovascular disease. There is also a wide range of drugs that have beneficial or harmful effects on plasma lipid concentrations. Studies aimed at detecting relationships between a set of drugs and movements of some classes of plasma lipid concentrations, of the importance of knowing whether the drug prescribed leads to non-compliance.

10. Evaluation Metrics in Predictive Analytics

In many research areas, data mining is increasingly reliant on predictive analytics aiming at developing accurate models, yet this predominately results in nontransparent "black-box" models without providing insights into the underlying data structure. The limited explanatory value of predictive models hinders their adoption and exploitation as decision support systems. In the critical care sector, intense research efforts have been made to translate the vast amount of data that is generated on a day-to-day basis into actionable knowledge aiming at improving the quality of care. A great number of publications deal with the use of intelligent methods in the analysis of Big Data in critical care. Among the various kinds of intelligent data analysis, predictive analytics is probably the most prominent. However, contemporary predictive analytics research in this field often results in models that are hard to understand and interpret by end-users. Models which do not explain why a particular prediction is made by highlighting altered patient conditions, treatment needs, contrary risk assessments, etc, impede the success of decision support systems for evidence-based medicine (EBM). Furthermore, recent reports also show that predictive models are often not validated effectively. Large-scale real-world validation on the performance capacity and generalization ability of predictive models is, as of today, rarely conducted in critical care research. There is a need for more detailed recommendations on how to validate predictive models in a clinical and critically ill environment [3]. A visual and freely available, Open Data Analysis Platform for the PIC is used to evaluate predictive models. With this revolutionary approach, decision makers such as physicians, medical staff, or even patients, access a visual representation and understand why and how predictions are generated.

10.1. Accuracy and Precision

Health statistics play an important part in transforming the results of health care and making it easier to process, assess, and treat the data. Precision health is a precise, customized, fortunetelling and preventive health care methodology. Such health practices predict, endorse, and prevent illness before they develop, in comparison to current health traditions of delivering care after an illness spying on symptoms. People can take advantage of such health predictions on time to prevent possible health threats. Health statistics as input statistics are fueling its ominous relational operation at a quick pace. Such health data approximately define a person and can be processed by clinicians and researchers in such health's opulent ecosystem. Health statistics, big facts, data mining, and artificial cognition are at the middle of the profound revolution of precision medicine [38]. They equip existing facts and numbers for the conduct of the intelligent circle analysis that generates a pivotal perspective into unstrained health. Such health advancements have the aptitude to innovate this time-honored care emphasis post illness with one absolute aptimony of pain patients.

Transforming the results of health care (HC) into software processing, evaluation, and handling health care data (HCD) and taking steps, e.g. a sequence of medical questionnaires, subsequent data obtained from a series of tests and the final response of the decision-making software. Each of them poses a challenge.

At a glance, the conclusions are vastly manifold, and a wide variety of academic interests have seldom been less harmonious, touching on many aspects of a patient's life and health-care system. Each patient has a number of unique health profiles that need to be interpreted in abundance, appealing to a wide range of academic realms. Patient life decisions and expectations are influenced by various factors outside the health-care system, such as diet, everyday life habits, or compliance with long-cycle drug administration, and are unique and exceptionally challenging in each case. In addition, on behalf of the institution, the eventual manner of making a determination is typically nonreproducible. Compound potential will, thus, be conferred the task of making an individual choice, which is fraught with systematic errors and beyond-study inference, affecting the validity of the final decision. Consequently, decisions made in this way on the basis of health care data can be properly interpreted in a variety of ways, are generally not reproducible, and are subject to speculation by many experts in decision-making.

10.2. Sensitivity and Specificity

It is common for a screening or diagnostic test to be evaluated by measuring its test performance characteristics. These are frequently summarized in terms of the test's sensitivity (the ability of the test to identify truly ill people), and its specificity (the ability of the test to identify truly normal people). Sensitivity and specificity will have a range of potential values from the minimum value when the ROC curve intersects the upper left corner cell to the maximum value when the ROC curve intersects the upper right corner cell [39]. It is common for such pairs to be reported at more than one threshold. Because, the test being evaluated could have different reasons for ordering the test under different conditions, the test could be processed using different equipment, or the underlying biology might dictate that the test will have better sensitivity than specificity, or vice versa, depending on the values of unobservable biological quantities. Use all available sensitivity and specificity pairs per study, synthesizing them in a multivariate framework that correctly accounts for the within-study dependence between the sensitivity and specificity pairs. Dukic and Gatsonis extended the hierarchical summary receiver operating characteristic (HSROC) model which represents as a bivariate, meta-analytic extension of the summary receiver operating characteristic (SROC) curve. Specifically, each alternative ROC curve is encoded in a latent space as a pair of equations linking the underlying true positive and false positive rates of the test. These pairs are extended to the multivariate case to accommodate studies where multiple sets of univariate sensitivity and specificity with their respective sampling variances are reported. For example, such a model can potentially allow the fixed-effect estimation of a unique SROC curve that represents a posterior non-ordered transformation of the bivariate curvilinear equations. As such, it is an extension of the originally proposed SROC model that allows for the only bivariate data. Two statistical models to analyze sensitivity and specificity data reported at more than one threshold.

11. Implementation Strategies for Predictive Analytics in Healthcare

On a daily basis, large volumes of health statistics are generated in the form of digitalized medical records, insurance claims, monitoring and case study data, as well as additional statistics that make predictions essential to abet decision-making in several healthcare areas. Internationally, researchers have used health statistics to create and validate predictive models with the purpose to support clinical decision-making, improve prevention, patient safety, public health, customer service, clinical research, or save costs. In the Dutch setting, validated predictive models are commonly applied to support the policy formation within healthcare planning. Furthermore, the patient-based predictive models are frequently used in the clinic to support decision-making regarding the prevention, diagnosis, selection of therapy, or patient monitoring. During the last decade, due to the increasing interest in transformation of healthcare outcomes into the customization of services (personalized medicine), academic research is carried out on predictive models varying from general statistical patterns to patient specificity.

From a theoretical perspective, relational health statistics offer an under-explored chance for the creation of a generic methodology to analyze and underpin the transformation of statistics into personalized analytics. These statistics are defined as those data that are directly or indirectly related to health and healthcare delivery and patient outcome, including clinical measurements, insurance data, government reports, data extracted from existing predictive models, statistics about treatment effectiveness, side effects, biographical or genealogic data. Personalized analytics are referred to as the automated approach to model and test predictive models by means of relational statistics, considering the specifics of the national healthcare system, available statistics, and the regulatory climate. This work concerns the development of an innovative methodology for testing personalized predictive models using relational health statistics [40]. The core understanding is embedded in design science research, and some failures which occurred during testing, as well as the insights gained therefrom, have served as a fundament for an iterative paradigm in the methodology development. The potential of applying personalized predictive models within the legal context of the public health professionalism field of practice is of particular interest. On the practical side, the country-specific challenges and focal points for predictive analytics in a quality dataset of Dutch relational health statistics are characterized for the first time.

11.1. Organizational Readiness Assessment

11.1. a Constructs

11.1. a.1 Negative side aspects

Scareceness of trained personnel argues that the number of required personnel is not sufficient.

Generally there exists the belief that the broader community was not prepared for a certain topic, had no business to be discussed or there was no need for them to be there. A corollary, which falls especially within the expectations of researchers, is that some of the presentations were "too academic".

One of the unpleasant aspects was the conflict between the chosen presentation and lunch break, or the fact that they fell into those presentations they were looking forward to the most.

Bad representation argues that the meeting was not appropriately planned and conducted in the representation of the municipality.

Unpleasant environment argues that the locality was not appropriate.

Time wasting included wasting of valuable time to get explanations by the lecturers. On the other hand, wastage of time in the opening of the series occurred through making references to inappropriate processes. On the third hand, the presenters refused to provide the answers to the questions which had been sent to them in advance.

Perception that presentations were disconnected is a belief that some people were not familiar with each other's activities, which was harmful to general understanding.

Incomprehensibility, i.e., the belief that some reports and data for themselves were not comprehensible, but that the altogether explanation was too obtuse.

Perceptual issues, i.e., that a certain number of sessions were held parallel, and at the same time, there was no sufficient time for proper consideration of the contents of each report.

Incompatibility with daily work argues that the matter did not fit within their daily work, which further had been sufficiently professional.

Requests for comments on certain topics, which, however, are much more the subject in the regional community.

Long duration positions concerns to the duration of the series, which was four days. Although the detailed program had been known in advance, to some extent such a long course had been imposed as a presumption.

Vague program, arguing that the relevant activities concerning the series were not perceived.

Concern about presence of a person that was considered out of the place.

Lack of published paper abstracts, including the belief that the abstract of each report should be before the beginning of the series, so that it could be given to the participants.

11.2. Change Management Approaches

The United States' healthcare industry abounds with domain knowledge. From the preferences and practices of individual doctors and patients to the broader trends in insurance claims and financial markets, a wealth of statistics is now accessible due to the continuously altered healthcare reform. By optimizing prevention and treatment regimens tailored to individual patients, national healthcare outcomes are transformed, and the delivery of personal medicine is revolutionized. Mathematic and statistic techniques, which are collectively forming a field known as data science, can harness this knowledge. Gathering, storing, retrieving, aggregating, manipulating, analyzing, modeling, and visualizing data be a broad array of purposes. Ensuring this is ethical, uses contemporary and scalable tools and is sustainable, as efficient and predictable development as possible, is provided by best software engineering and database design practices. As illustrated by a case-study of a predictive model for A&E revisit, the process, insights, and principles of data science, strategic planning, and healthcare are demonstrated.

The challenges of implementing change, frequently underpinned by health statistics, are well-documented. The multifaceted interventions needed to bring primary care in line with medical home standards is seldom explored from a political perspective, with little research on how practitioner beliefs affect subsequent reorientation. For a case-study of six extensively regulated primary care practices, content analysis is used to explore how practitioners used, managed, and contested four state-driven performance metrics. Similar and disciplined contrasts illustrate whether beliefs change in light of performance data. It is found that earlier beliefs may constrain the range of reactions to data, and that strong beliefs may be fueled by knowledge silos resistant to external data sources. Such insights have implications for driving successful reform, suggesting that policy makers need to conceive of redesign as a process that extends over multiple years, working to generate engaged ownership among medical staff.

12. Collaborative Partnerships in Healthcare Analytics

Recently, there has been an exponential rise of data within the healthcare sector, including electronic health records (EHRs), due to affordability of electronic data sources, which can be exploited for improving patient care. Organizations are taking up the challenges of operations optimization have been shifting their focus to leveraging historical data, not only to review and adjust their processes but also to predict system inputs or outcomes. In this context, several statistical techniques and machine learning methodologies, with the overall term predictive analytics (PA), have been developed to exploit the information believed to be extracted from data. This interest has generated a new era in the use of large-scale datasets for improvement of healthcare outcomes and introducing customizable care, the so-called precision medicine, contributing towards reducing costs and enhancing the quality of services.

Nevertheless, the inherently multidisciplinary nature of efficient PA implementation in healthcare is still posing some challenges. The HarmonicSS project as an example has enforced a transnational network of collaborating healthcare institutions to create a powerful platform on top of EHRs, able to offer personalized predictions of disease risks and the related early interventions [41]. As healthcare is a sensitive domain, fostering wide adoption of specific data-driven solutions is a complex process, depending on regulatory, ethical, and social aspects.

12.1. Academic-Industry Collaborations

Academic-industry collaborations are proving to be effective in translating population health data, traditional and digital health methodologies into predictive analytics decision models [42]. These predictive analytics decision models may be the first of a suite of solutions capable of systematically addressing challenges and exploiting the infrastructures for personalized medicine for diverse populations. Academic-industry collaboration was tested as a care model for group-based psycho-educational sessions employing the evidence-based materials of an Internet utilized, decision support for persons concerned about a loved one with suicide risk. Academia is a key player in the development of predictive analytics and population health frameworks and essential to the progress of necessary policies governing their development. Academia is actively involved in the resolution of the data interoperability and handoff issues with traditional clinical and additional data sources to effectively translate population health data, methodologies and technologies addressing the underlying psychosocial issues associated with population health and differential care into decision models useful at the point of need [41]. Academic-industry collaborations should be enhanced and carefully negotiated to balance the needed research, technology and care developments. A "best practices" "decision point and care flows" model for academic-industry collaborations would encompass four stages of activities through 14 steps which begin with the mutually agreed upon policy guidance development of document and research priority settings.

12.2. Public-Private Partnerships

Public–Private Partnerships (PPPs) have long been leveraged worldwide to advance various goals in healthcare. Given the data-intensive nature of predictive analytics, PPPs may hold significant and unique promise for both advancing health outcomes and addressing the "data divide" [43]. Partnering diverse stakeholders with complementary areas of expertise allows for targeted focus across the successive layers of the data, analytic methods, and health expertise on the same datasets and problems. Harvesting big data for precision health from cloud EHRs and aggregating it with the myriad individual-level datasets available requires savvy data deals and funded infrastructure, often controlled by industry or government. "Big data" are often only as clean (i.e., coded, standardized, complete) and population-representative as they are made to be, which can obfuscate pre-trained models and correlations in big data that are neither clinically relevant nor causative—this black box issue becomes essentially a matter of trusts and policy control. Inpatient care and associated EHRs account for only about 35% of healthcare spending in the US, and only about 20% of the time in a year for individuals occurs in a healthcare setting. Positive health outcomes are influenced by a cornucopia of upstream social determinants, yet EHR data infrastructures are tightly contained within proprietary silos that incentivize the monolithic use of tools towards a narrow band of profitable health outcomes. Per PPPs, many of the requisite components for analytic predictive models of health outcomes sit outside academic arenas—proprietary algorithms, rich EHR data troves for classes of health outcomes, and domain expertise of how broad data deals were brokered [41]. However, strong ethical, equity, and stewardship concerns also arise when ostensibly public data are commercialized through secretive partnerships, particularly when the data deal is more about "cloud-wash" than depersonalizing and integrating EHR—forcing a more concrete awareness and substantial role for the public good beyond legal boundarization.

13. Emerging Trends and Innovations in Health Statistics

Developments in statistical analysis and related innovations are pivotal to the future of public health practice. This chapter identifies and explains emerging trends and innovations in health statistics. The chapter catalogues emerging key areas of importance to public health and enumerates numerous examples at the forefront of innovative statistical practice in public health settings. Lastly, the chapter offers global perspectives on the future of health statistics.

Traditionally, health statistics have played a vital role in public health and as evinced by [32] health statistics have a long and storied tradition in public health. Most public health advances in the past century have been predicated on robust statistical analysis informed by high quality data. Despite this, the field is not static and advances in statistics, as well as shifts in funding and research priorities, continue to shape how health data are analyzed and used. Importantly, developments in technology and data engineering affect how health data are collected and stored, and these changes have implications for analysis and reuse of these data.

13.1. Blockchain Technology in Healthcare

Blockchain technology, a means of securely distributing an encrypted digital ledger to a network of participants, is recommended for secure handling of medical data. Remarkable development is anticipated in remote patient monitoring, big data analytics of patient health records and patient behavior data, and preventive healthcare applications in smart hospitals. A new umbrella concept of the HealthcareThingsSpace integrates smart healthcare devices and applications with healthcare systems in hospital to better utilize their data and make more informed decisions in order to improve patient treatment and well-being [44]. Still, healthcare smart devices are underdeveloped in terms of secure data exchange and processing. A promising approach is to connect these devices with the blockchain technology, which is a decentralized platform storing data in linked encrypted ledgers that are immutable and unchangeable. Login systems were developed based on access and intelligent contracts to ensure secure data exchange between the system in hospital and patient healthcare smart devices. The concept is developed with regard to the blockchain based implementation ensuring secure data exchange not only of the sensor acquired patient vital signs data with the HealthcareThingsSpace system in hospital but also healthcare recommendations and instructions with regard to the patient treatment back to the patient. Secure monitoring IOS and Android applications for hospital staff and patients were developed and verified in laboratory environment. Secure data processing steps of the Acquired Patient Vital Signs Data by the Blockchain Enabled HealthcareThingsSpace System in Hospital include the Consensus Mechanism, Secure Ledger of Patient Vital Signs, Smart Contract Execution, and Data Storage in the HealthcareThingsSpace.

13.2. Internet of Medical Things (IoMT)

The services and the facilities in the healthcare industry are prospering and enhancing at a significant pace. The developments and the advancement in the medical industry are sweating to improve the human health condition and to diagnose the disease efficiently and precisely. The health-tracking devices such as the digital variety of watches and wearable devices are accelerating the potential of predicting the disease conditions based on the physiological parameters. The variety of innovative mechanisms and initiatives are conjoined with the health discipline that are fostering security and enhancing the quality of healthcare. Few such strategic techniques and domains that are toil to improve the patient condition or are complicate to effect health condition in many fields are prospered to analyze predominate scenarios. One such crucial and prominent scenario that is being adequately effective to the cyber health industry which is reported due to infected physiologic states or disturb physiological states. The infected states can be ailments or signs or symptoms that are tend to a specific malady or disease condition. This is prosperous, if detected well in advance, then the perturbation or damage can likely to prevent.

14. Conclusion

Predictive analytics are analytics that can be used to predict future outcomes using historical data. Most predictive analytics models enhance critical planning, decision-making, and optimization activities and are used to describe patterns, and to identify associations. In healthcare, clinicians use predictive analytics to guess what will happen in the future. Predictive analytics uses current predictive patterns to predict what will happen in the future. It is a cost-saving activity and determines which key factors would lead to the risky event. The data can be patient or geographical datasets, irrespective of their nature. Patient clinical identification, understanding the user profile, and fraud mapping in the health claims are some of the common applications of this activity, leading to significant data changes. Patterns can be detected across all these applications by ingesting and processing a healthcare dataset. Pre-processing and exploratory visualization will make the data useful. Challenges, optimized computing, the optimized model, the evaluation, and the implementation of strategy, will help in translating patient patterns. Dimensions can vary depending on the analysis, but models such as scores, odds, likelihood, and tree, future propensity, classification or output client, can transform the health sector. The four critical components of predictive analytics in healthcare are essential, and realizing the solutions, governance guidance, and reliability measurement will carry significant financial and valuable rewards to the patient.

References:

1. A. L. MUTAIRI ALYA O., "The Association Between Statistical Sciences And Medical Science Based On Previous Studies," 2014. [PDF]

2. J. Ni, B. Chen, N. M. Allinson, and X. Ye, "A hybrid model for predicting human physical activity status from lifelogging data," 2019. [PDF]
3. S. Van Poucke, Z. Zhang, M. Schmitz, M. Vukicevic et al., "Scalable Predictive Analysis in Critically Ill Patients Using a Visual Open Data Analysis Platform," 2016. ncbi.nlm.nih.gov
4. R. Osuala, J. Li, and O. Arandelovic, "Bringing modern machine learning into clinical practice through the use of intuitive visualization and human-computer interaction," 2019. [PDF]
5. R. Chen and I. Paschalidis, "Learning Optimal Personalized Treatment Rules Using Robust Regression Informed K-NN," 2018. [PDF]
6. R. J. DeRubeis, Z. D. Cohen, N. R. Forand, J. C. Fournier et al., "The Personalized Advantage Index: Translating Research on Prediction into Individualized Treatment Recommendations. A Demonstration," 2014. ncbi.nlm.nih.gov
7. C. H. Huang, F. A. Batarseh, A. Boueiz, A. Kulkarni et al., "Measuring Outcomes in Healthcare Economics using Artificial Intelligence: with Application to Resource Management," 2021. [PDF]
8. C. Scott Kruse, R. Goswamy, Y. Raval, and S. Marawi, "Challenges and Opportunities of Big Data in Health Care: A Systematic Review," 2016. ncbi.nlm.nih.gov
9. C. C. Yang, "Explainable artificial intelligence for predictive modeling in healthcare," Journal of healthcare informatics research, 2022. springer.com
10. K. Batko and A. Ślęzak, "The use of Big Data Analytics in healthcare," Journal of big Data, 2022. springer.com
11. M. Hassan, F. M. Awan, and A. Naz, "Innovations in genomics and big data analytics for personalized medicine and health care: a review," International Journal of …, 2022. mdpi.com
12. S. Sharma, K. Chen, and A. Sheth, "Towards Practical Privacy-Preserving Analytics for IoT and Cloud Based Healthcare Systems," 2018. [PDF]
13. G. Gürsoy, M. Doerr, J. Wilbanks, J. K. Wagner et al., "Navigating ethical quandaries with the privacy dilemma of biomedical datasets," 2020. ncbi.nlm.nih.gov
14. D. Hartzband and F. Jacobs, "Deployment of Analytics into the Healthcare Safety Net: Lessons Learned," 2016. ncbi.nlm.nih.gov
15. L. Prudente Moorman, "Principles for Real-World Implementation of Bedside Predictive Analytics Monitoring," 2021. ncbi.nlm.nih.gov
16. D. Dixon, H. Sattar, N. Moros, S. Reddy Kesireddy et al., "Unveiling the Influence of AI Predictive Analytics on Patient Outcomes: A Comprehensive Narrative Review," 2024. ncbi.nlm.nih.gov
17. …, S. Bischoff, D. Samhammer, K. Seitz, and P. A. Fasching, "Perspectives of patients and clinicians on big data and AI in health: a comparative empirical investigation," AI & SOCIETY, 2024. springer.com
18. A. Majeed and S. O. Hwang, "Data-driven analytics leveraging artificial intelligence in the era of COVID-19: an insightful review of recent developments," Symmetry, 2021. mdpi.com
19. V. Kumar, C. Prabha, and M. M. Hasan, "Unlocking Gender-Based Health Insights with Predictive Analytics," in … -Based Healthcare with AI and …, 2024. [HTML]
20. W. H. Weng, "Machine Learning for Clinical Predictive Analytics," 2019. [PDF]
21. H. Habehh and S. Gohel, "Machine Learning in Healthcare," 2021. ncbi.nlm.nih.gov

22. M. Shehab, L. Abualigah, and Q. Shambour, "Machine learning in medical applications: A review of state-of-the-art methods," Computers in Biology, Elsevier, 2022. uts.edu.au
23. G. Quer, R. Arnaout, M. Henne, and R. Arnaout, "Machine Learning and the Future of Cardiovascular Care: JACC State-of-the-Art Review," *Journal of the American College of Cardiology*, 2021. jacc.org
24. C. Bhatt, I. Kumar, V. Vijayakumar, and K. U. Singh, "The state of the art of deep learning models in medical science and their challenges," Multimedia Systems, 2021. researchgate.net
25. G. Z. Papadakis, A. H. Karantanas, M. Tsiknakis, A. Tsatsakis et al., "Deep learning opens new horizons in personalized medicine," 2019. ncbi.nlm.nih.gov
26. C. W L Ho, J. Ali, and K. Caals, "Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance," 2020. ncbi.nlm.nih.gov
27. J. K. Paulus and D. M. Kent, "Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities," 2020. ncbi.nlm.nih.gov
28. A. Alyass, M. Turcotte, and D. Meyre, "From big data analysis to personalized medicine for all: challenges and opportunities," 2015. ncbi.nlm.nih.gov
29. Z. Ahmed, "Practicing precision medicine with intelligently integrative clinical and multi-omics data analysis," 2020. ncbi.nlm.nih.gov
30. L. Möhrmann, M. Werner, M. Oleś, and A. Mock, "Comprehensive genomic and epigenomic analysis in cancer of unknown primary guides molecularly-informed therapies despite heterogeneity," Nature, 2022. nature.com
31. M. Pierobon, N. J. Robert, D. W. Northfelt, "Multi-omic molecular profiling guide's efficacious treatment selection in refractory metastatic breast cancer: a prospective phase II clinical trial," Molecular, 2022. wiley.com
32. L. Chen, "Healthcare data mining from multi-source data," 2017. [PDF]
33. P. Berchialla, C. Lanera, V. Sciannameo, D. Gregori et al., "Prediction of treatment outcome in clinical trials under a personalized medicine perspective," 2022. ncbi.nlm.nih.gov
34. B. A. Williams, "Constructing Epidemiologic Cohorts from Electronic Health Record Data," 2021. ncbi.nlm.nih.gov
35. W. Yeon Park, K. Jeon, T. Sippel Schmidt, H. Kondylakis et al., "Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension," 2024. ncbi.nlm.nih.gov
36. M. Iwagami and H. Matsui, "Introduction to Clinical Prediction Models," 2022. ncbi.nlm.nih.gov
37. D. Bzdok, D. Engemann, and B. Thirion, "Inference and Prediction Diverge in Biomedicine," 2020. ncbi.nlm.nih.gov
38. C. Thapa and S. Camtepe, "Precision Health Data: Requirements, Challenges and Existing Techniques for Data Security and Privacy," 2020. [PDF]
39. T. A. Trikalinos, C. M. Balion, C. I. Coleman, L. Griffith et al., "Chapter 8: Meta-analysis of Test Performance When There is a “Gold Standard”," 2012. ncbi.nlm.nih.gov
40. S. Creemers, "Determinants for successful deployment of clinical prediction models : a design science research in the Dutch healthcare sector," 2019. [PDF]
41. A. C. Arevian, D. Bell, M. Kretzman, C. Kasari et al., "Participatory methods to support team science development for predictive analytics in health," 2018. ncbi.nlm.nih.gov

42. D. Alemayehu and M. L. Berger, "Big Data: transforming drug development and health policy decision making," 2016. ncbi.nlm.nih.gov
43. P. Granados Moreno, Y. Joly, and B. Maria Knoppers, "Public–Private Partnerships in Cloud-Computing Services in the Context of Genomic Research," 2017. ncbi.nlm.nih.gov
44. X. Zhang, R. Sharma, and S. C. Wingreen, "Block-chaining in precision healthcare: a design research approach," 2018. [PDF]ss