

Critical Evaluation of the Impact of Big Data on the Drug Development Stage: A Literature Review

Mahmoud Mansi*

Department of Life Sciences and Education, Staffordshire University,
Stoke-On-Trent, ST4 2DE, United Kingdom

Review Article

Received: 03/01/2022, Manuscript No. JPPS-22-51721; **Editor assigned:** 10/01/2022, Pre QC No. JPPS-22-51721 (PQ); **Reviewed:** 24/01/2022, QC No. JPPS-22-51721; **Revised:** 30/01/2022, Manuscript No. JPPS-22-51721 (A); **Published:** 08/02/2022, DOI: 10.4172/2320-1215.11.1.006

For Correspondence:

Mahmoud Mansi, Department of Life Sciences and Education, Staffordshire University, Stoke-On-Trent, ST4 2DE, United Kingdom

E-mail:

mahmoud.mansi@research.staffs.ac.uk

Keywords: Big data, Drug discovery, Target discovery, Drug development, Disease

ABSTRACT

Drug production has become an expensive and time-consuming procedure with an incredibly poor performance rate and a failure to account for human differences in drug reaction and toxicity. Throughout the last decade, an emerging 'big data' method that is focused on the advancement of electronic resources of chemical compounds, disorder genotype markers, operational outputs, and clinical knowledge concerning cross genetic anomalies and toxic effects has grown at an exponential rate. This paradigm transition has allowed the systemic, high-throughput and rapid detection of new drugs or recycled indicators of established drugs for infective molecular anomalies that are unique to each patient. The growing involvement in big data from the digital technology world and interactive genetic testing sectors has made it easier to obtain customised, precision medicine. Assurance (QA) is critical in the pharmaceutical sector for ensuring that pharmaceutical goods are prepared to a safe and uniform standard. QA is a broad term that refers to anything that can affect the quality of a drug during its research, development, manufacturing, and distribution phases. QA specialists are in charge of implementing a variety of methods that help to ensure the quality of a medicine.

INTRODUCTION

The advancements in bioinformatics, sequencing, and data processing technology have resulted in the generation of massive amounts of complex data available for drug development. It is becoming increasingly popular to analyse these databases in order to further investigate and explain illness and find innovative medicines. Current open data projects in fundamental and clinical science have greatly expanded the forms of data that are now accessible to the public. Big Data (BD) has been successfully used in several fields, including the drug development process, during the last few years [1].

In this critical evaluation, the extent to which the drug development process has been transformed by BD will be explored.

Drug development has typically proven to be a lengthy and expensive multi-step operation. Conventionally, potential drug targets have been identified and evaluated using a reductive method or a Closed-World Assumption (CWA), which is focused on a restricted interpretation of biology and is limited to modifying one molecular mechanism. Due to our poor knowledge of systems biology, each stage in drug research and production has been fraught with complexity, culminating in an incredibly poor success frequency. A modern medicine requires huge amounts of money in funding and an estimate of 9–12 years to reach the industry [2].

Existing drug development difficulties involve; (i) a restricted capacity to adequately define and/or track molecular processes of relevance; (ii) a lack of sufficient laboratory mechanisms to evaluate product drugs/perturbations and their therapeutic potential, and; (iii) a number of associated numerous off-target consequences of the compounds that are relatively overlooked (can be alluded to as “polypharmacology” [3]. Even though there was an increased research and development investment, the rising incidence of depletion in the latter periods of drug development highlights the necessity for novel alternatives to drug development [4].

LITERATURE REVIEW

Big data

Until recent times, research advanced on the basis of small data sets that were generated in closely regulated formats utilising testing strategies that limited their scale, materiality, and volume, and were very rigid in their management and development. Although several of these small data are quite significant, they lack other features of BD. National censuses, for instance, are usually created once a decade and pose about 30 formalised issues once conducted, it is difficult to change or insert queries. BD, on the other hand, is produced constantly and is more versatile and efficient in its processing [5].

Alternatively, instead of dwelling on the ontological features of what comprises the essence of BD, others describe it in terms of the technical complexities involved in collecting and analysing it, or in maintaining it on a single computer. BD tests traditional analytical and simulation methods and pushes the boundaries of computing capacity to analyse them [6].

The term ‘big data’ usually applies to a method of collecting quantitative data in an impartial way, excluding previous assumptions, and then analysing it using data processing techniques to generate new concepts. The study of nucleic acid and enzyme genomes collected in publicly available data, along with DNA microarray-based genetic code transcriptome and DNA molecular variability results, has contributed to this methodology in cell genetics [7].

With exponential rates of advancement in genetic analysis techniques and significant progress in international data management and distribution technology, the big data method has become enhanced and incorporated with various forms of data including epigenomic attributes, bio-ontologies, structural characteristics, membrane proteins, electronic health records, clinical research registrations, and clinical safety. Simultaneously, data collection algorithms and technologies designed specifically for this form of data were extensively built [8].

The adaptation of computational methods to the increased size and sophistication of databases has become a significant challenge in large data research. Adjustment of p-values for Multiple Hypothesis Testing (MHT), for example, is a critical problem in attempt to monitor erroneous exploration when eliminating inaccurate negatives. Furthermore, high-dimensional information frequently necessitates dimensionality elimination through data simulation strategies including vital element removal, non-negative neural networks, and molecule mechanism refinement of the data. Centred on genetic and medical evidence, several unmonitored and controlled machine learning algorithms have been applied to the study of biological big data to classify unidentified infection subgroups, clarify new disease goals, and forecast therapeutic results [9].

Big data available for drug discovery

The recognition and interpretation of disease mechanisms accompanied by target detection is always the first step in drug development and contributes to drug discovery. One condition identification pattern in drug development is the transition from a symptom-driven disorder identification method to a precise medicinal practice based on molecular elements like predictive diagnostics. This leads to more effective screening [10]. Creating a new disease classification necessitates the molecular markers of all disorders. Furthermore, an optimal degree of disease awareness will include all molecular modifications, from DNA to RNA to proteins, and the impact of external causes [11].

At the DNA stage, one form of DNA sequencing variance commonly employed to classify disorder is Single-Nucleotide Polymorphisms (SNPs) that arises uniquely in the patient group. Copy Number Variations (CNVs) represent comparatively broad areas of gene mutations that can be linked to illness. SNPs and CNVs can be found

using Genome-Wide Association Studies (GWASs) and entire decoding. Variations, especially somatic genetic changes, are extensively investigated through future generational screening in malignancy to identify genetic variants that impart a preferential development benefit to cells [12].

Gene expression (predominantly mRNA) is potentially the more commonly utilised element for disorder classification at the RNA stage. Due to the advancement of microarray science, it has become widely adopted to better explain disease mechanisms. The latest advancement of RNA-Sequence shows promise in terms of increased transcription availability and identification of low abundance datasets. RNA-Sequence has developed successfully to research tumour and host associations and it is also being used to examine Neurological Disorders (ND) and neurocognitive illnesses. Furthermore, it is proving to be a highly effective method for studying quantitative phenotype loci correlated with expression of genes in complicated disorders [13].

An overview of the structural shifts in pathology can now be easily simulated using a collection of datasets obtained using various methods. The latest advancement in single cell processing introduces a new dimension of genetic improvements. As we understand the complex mechanism of clinical development, the number of levels rises significantly. Furthermore, in addition to disease specimens from individuals, various preclinical constructs (e.g., cell lines and animal experiments) may be molecularly classified to further explain illness and test hypotheses [14].

In the pharmaceutical aspect, structural modifications in model systems perplexed by biochemical or biological factors may be recorded in order to further explain disorder and medication mechanisms. Genetic code adaptive displays, including such RNAi and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9, may be used to investigate gene expression and genetic regulation systems [15]. Furthermore, large data sequencing will easily identify the immune functions of hundreds of organic molecules in a wide variety of disorder types. Due to the accessibility of Electronic Medical Records (EMRs) and experimental testing, patient responses to medication action can now be monitored and analysed. As well as the biological and biochemical evidence, free-text data from the research may help with drug development [16].

Big data sources

No one facility, organisation, or collaboration may provide data that captures all of the complexities of dynamic disease environments. Furthermore, recognising these processes requires a significant amount of data, so predictive capacity can be achieved. To comprehend disease and find new therapies, interdisciplinary examination of several levels of datasets from various sources is needed. As a result, it is important that the details be available to the public in order to allow for any and all knowledge to be conveniently linked. Several valuable research datasets that can be utilised for drug research have lately been developed and published. Some of these examples include, COSMIC, GEO, Human Protein Atlas and PubChem [17]. Public databases have not just been extensively utilised as a point of comparison, but have also often been rigorously analysed to raise additional issues, discover novel observations, and also test hypotheses [18].

Target discovery

Utilising big data to identify targets for clinical trials frequently begins with identifying molecular differences among disease specimens. The chemical variations are associated with genetic alterations, genetic variance, or other characteristics, and they are often employed to guide target development. For example, UK Biobank is a huge genetic archive and study platform that contains detailed genetic and wellbeing records from over 500,000 UK individuals. The database is constantly updated with new information and is available to licensed researchers worldwide who are undertaking critical investigations into increasingly serious and life-threatening conditions. It has made significant contributions to the development of medical health and therapy, as well as many technological advances that have improved public health [19].

Our exponentially evolving capacity to classify organic compounds at the whole-genome stage has resulted in a conceptual change in the area of medicinal drug discovery. Throughout the last period, the study of genetic DNA degenerative changes inside close protein-coding regions, and transcriptome profiling, have resulted in the creation of innovative methods to allow big information, unbiased objective discovery. We may therefore calculate genome-wide hypermethylation, chromatin protein changes, splicing mutations, transcription factor binding locations and protein accumulation. This technical advancement also culminated in ever-expanding large data warehouses for biomedical science and drug production, allowing for hypothesis-free, unbiased goal exploration [20].

Many cancer genomic studies have generated a database of somatic DNA structural modifications including single nucleotide changes, minor deletions, copy number modifications, and genetic translocations within every cancer class as possible cancer triggers and novel clinical goals. Comprehensive collections of transcriptional control mechanisms, like genome-wide transcription factor binding locations for significant transcriptional regulators, along with epigenetics markers like histone protein acetylation, have recently been established to aid in the identification

of clinically active drug candidates [21]. Transcriptome repositories are being widely used. These archives such as the mentioned above, in conjunction with the increasing proliferation of extremely specific chemical collections, have significantly reduced the possible time needed from drug discovery to therapeutic implementation.

Outstanding challenges

Assessments taken from disease specimens can be of low standard. According to recent research, a substantial percentage of tumour cultures are impure due to the presence of combined resistant and stromal cells. Moreover, there is a wide range of technological and biological variability across samples. Furthermore, the consistency of materials, particularly antibodies, differs significantly [22]. Misappropriation of antibodies can ultimately result in study failure. Finally, while the database from high throughput studies is valuable as a comparison method for detecting expression or even as a method for inferring biological activity, it also produces erroneous signals culminating in the misinterpretation of possibly successful targets [23].

Amongst the most difficult tasks for scientists is integrating several levels of data into a functional and structured framework for further understanding, drug discovery and patient treatment. Integration of various “omics” evidence with medical physiology knowledge reported in EMR is critical for identifying clinically specific pathogenic molecular modifications as drug targets. Topology-related patient-patient networks, for example, based on integrative clinical and genomic evidence from 11,000 people could classify three new subgroups of type 2 diabetes. Incorporation of complex genomic and phenomic content, diagnostic data and tests, while also environmental and social factors, when adequately utilised and controlled, would eventually revitalise clinical practice [17].

Disease biomarkers and the issues that arise with their use

A major difficulty that is often found is the lack of proper understanding we have for nervous system disorders. Austin has indicated that if these disorders and their mechanisms were comprehended better, more optimal treatment could be developed. It is imperative that clinical observations throughout the development phase are noted precisely so that we can benefit [24]. The issue remains that developing biomarkers for diseases require us to initially understand the inherent biological mechanisms of the disease in order to produce the necessary treatment and drugs. A novel method to distinguish types of disease, based on biomedical big data, is Big-data-based edge biomarker. This is done through a large network, allowing for different approaches and tactics to specify disease in more individual samples [25].

Genomic big data and biomarkers

Clinical samples can allow us to compare genomic profiles in order to identify biomarkers. Some examples comprise the discovery of EGFR mutations to test gefitinib sensitivity. Another includes the detection of a 12-gene colon cancer sequence in order to predict potential recurrence in patients treated with Leucovorin and Fluorouracil. An important piece of research entails, where quantitative reverse transcription polymerase chain reaction of 375 genes was done in patients with colon cancer [26]. In this study, patients were observed for a period of three years following treatment with surgery, or both surgery and Fluorouracil or Leucovorin. It was possible to identify 48 genes linked to a higher recurrence risk and 66 genes showed benefit from the drugs given. From the 66, 7 genes were highlighted due to their biological basis and recurrence links and compared to 5 reference genes which resulted in the creation of a recurrence score to calculate recurrence risk.

Unresolved challenges with disease biomarkers

There is a vicious cycle with biomarkers in that the deficiency of these biomarkers can result in clinical trial failure, but clinical trials allow for the primary discovery of these essential biomarkers. Furthermore, clinical trial complexity and disparity can allow biomarkers to be overlooked. A solution is performing an integrative analysis of trials within a number of studies. However, many trials are currently unavailable to the public, although shared data is crucial in detecting successful biomarkers for interventions or understanding why drugs fail through identifying correct targets and their demographics [27].

Combinatory treatments

Combinatory treatment strategies were initially introduced in the mid-twentieth century in the form of cancer treatment to improve efficacy and reduce adverse effects of anti-cancer treatment. Nevertheless, with a greater emphasis on managing diseases with multifactorial pathophysiology, such as hypertension, diabetes, and heart disease, it is increasingly clear that the ‘one medication-one goal’ strategy might be oversimplified, and that hybrid pharmacological treatments should be reconsidered in a wider context than just cancer therapy. Furthermore, with the advent of contemporary genomics and a systems-medicine method, the traditional “mechanism of action” has given way to a wider “signature”-based predictions, providing significant insight into drug method [28].

Combinatory treatment's possible pathways of effect involve compatible acts, in which two or several medications attack different receptors within the same protein or channel, anticounter active behaviour, in which a drug inhibits the pharmacological reaction to a first drug, and promote actions in which the second drug stimulates the function of the first drug. Identifying several antibody targets in melanoma therapies is a new promising illustration of combinational drug layout. Examples of combinatory treatments are found in breast cancer, hypertension and Alzheimer's disease. For melanoma, the grouping of a B-Raf enzyme inhibitor and an extracellular signal-regulated inhibitor has been suggested [29].

Techniques for theoretical evaluation of combinational drug therapy are still being established, but they involve analytical model-based methods, such as network-based processes focusing on the principle of gene locality, or biological model-based approaches in situations in which the pathophysiology of a disorder mechanism is well known. Given a greater comprehension of these methods, a number of problems remain. This technique has been hampered in general from the inadequate physiological knowledge of the diverse biological processes concerned, as well as the heightened risk of toxicity associated with the use of combinational pharmacological treatment [30].

Outlook

One existing drug development theory assumption is that fully studying infection molecular modifications can eventually contribute to the development of new therapeutics drugs. To identify molecular modifications in disorders and respond to drug treatments, molecular profiles must be viewed in an open manner, that we now refer to as 'big data'. Due to the rapid technological advancements, there is no question that the identities we have developed would soon become limited collections. Much broader and more nuanced databases would be developed in the foreseeable future to classify health processes: from single cells to tissues or tumour cells to microbes. The enormous BD datasets signals a once-in-a-lifetime chance to use them to accelerate research right now [31].

Considering the scale and sophistication of drug discovery databases, no one individual or group may apprehend or utilise everything; thus, the whole drug discovery process must be re-engineered, with data and robust data simulations driving every phase. Suggestions of measures involve selecting suitable blood samples to evaluate, selecting suitable templates to test hypotheses, and so on. Furthermore, although high performance computation enables us to produce theories rapidly, the existing laboratory conditions restrict our verification attempts [10].

Future potential

The Information Technology (IT) sector's increasing engagement and participation is changing the sector by offering platforms for data collection, exchange, and research, often without the participation of conventional medical organisations, through portable technologies specifically linked to cloud-based market storage/analysis centres. For example, as part of the Precision Medicine Initiative, the National Institution of Health is attempting to establish a quantitative cohort of one million or more Americans, with comprehensive analysis of pharmacological samples and advanced susceptibility evaluation through smart phones or monitoring devices. The combination of this information with the plethora of other data sets provided for each person provides future hope in reaching the target of precision medicine [32-34].

CONCLUSION

Big data drug discovery has grown in prevalence and speed over the last few decades. The research began with proof-of-concept trials, then expanded to include implementation of innovative techniques covering various disciplines, the collection of experimental results for general application and research, and is now expanding to include more complex biological aspects and therapeutic settings for greater validity of real-world drug production problems. Big data drug discovery can make it easier to identify treatment solutions for unusual subgroups of prevalent disorders, rare emerging disorders and health problems that affect biologically insignificant demographics.

REFERENCES

1. Pastorino R, et al. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. *Eur J Public Health*. 2019; 29(Supplement_3): 23–27.
2. Mohs R, et al. Drug discovery and development: Role of basic biological research. *Alzheimer's & dementia* (New York, N Y.). 2017; 3(4): 651–657.

3. Yildirim O, et al. Opportunities and Challenges for Drug Development: Public-Private Partnerships, Adaptive Designs and Big Data. *Front pharmacol.*2016; 7: 461.
4. Spadoni C, et al. Pediatric Drug Development: Challenges and Opportunities. *Curr Ther Res Clin Exp* 2018; 90: 119–122.
5. Kitchin R, et al. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *J Big Data.* 2016.
6. Kibira D, et al. Integrating Data Analytics and Simulation Methods to Support Manufacturing Decision Making. *Wint sim conf.* 2015; 2100–2111.
7. Quiñones RM, et al. Climate change vulnerability of freshwater fishes in the San Francisco Bay Area. *San Fran Estu Wat Sci.* 2017; 18(2): 412.
8. Pastorino R, et al. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. *Eur J Public Health.*2019; 29(Supplement_3): 23–27.
9. Atwell K, et al. How computational models contribute to our understanding of the germ line. *Mol Reprod Dev.* 2016; 83(11): 944–957.
10. Qian T, et al. Use of big data in drug development for precision medicine: an update. *Expert Rev Precis Med Drug Dev* 2019; 4(3): 189-200.
11. Liu B, et al. Artificial intelligence and big data facilitated targeted drug discovery. *Str vas neu.*2019; 4(4): 206–213.
12. Zarrei M, et al. A large data resource of genomic copy number variation across neurodevelopmental disorders. *Npj. Genom Med.*2019; 4:26.
13. Lähnemann D, et al. Eleven grand challenges in single-cell data science. *Genome Biol.* 2020; 21: 31.
14. Angerer P, et al. Single cells make big data: New challenges and opportunities in transcriptomics. *Curr Opin Syst Biol.* 2017; 4: 85-91.
15. Kato-Inui T, et al. Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 with improved proof-reading enhances homology-directed repair. *Nucleic Acids Res.* 2018; 46(9):4677–4688.

16. Agrawal R, et al. Big data in digital healthcare: lessons learnt and recommendations for general practice. *Hum.*2020; 124: 525–534.
17. Chen B. Leveraging big data to transform target selection and drug discovery. *Clin. Pharm. Therap.*2016; 99(3):285-297.
18. Khoury M. J, et al. The intersection of genomics and big data with public health: Opportunities for precision public health. *PLoS Med.* 2020; 17(10):e1003373.
19. Wu Y, et al. Genome-wide association study of medication-use and associated disease in the UK. *Biobank Nat com.*2019; 10(1): 1891.
20. Glicksberg B, et al. Leveraging Big Data to Transform Drug Discovery. *J Mol bio.*2013; 138.
21. Dlamini Z, et al. Artificial intelligence (AI) and big data in cancer and precision oncology. *Comput Struct Biotechnol.*2020; 18: 2300–2311.
22. Nakamura H, et al. Organoid culture containing cancer cells and stromal cells reveals that podoplanin-positive cancer-associated fibroblasts enhance proliferation of lung cancer cells. *Lung cancer (Amsterdam, Netherlands).*2019;134:100–107.

23. Paul D, et al. Artificial intelligence in drug discovery and development. *Drug Discov.* 2021; 26(1), 80–93.
24. Mayeux R. Biomarkers: potential uses and limitations. *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics.* 2004;1(2),182–188.
25. Frangogiannis N G, et al. Biomarkers: hopes and challenges in the path from discovery to clinical practice. *Translational research: the journal of laboratory and clinical medicine.* 2012; 159(4): 197–204.
26. O'Connell MJ, et al. Relationship between tumor gene expression and recurrence in four independent studies of patients with stage II/III colon cancer treated with surgery alone or surgery plus adjuvant fluorouracil plus leucovorin. *Clin Oncol Am Clin oncol.* 2010; 28(25): 3937–3944.
27. Wilson JL, et al. Biomarkers: Delivering on the expectation of molecularly driven, quantitative health. *Exp. Biol. Med. (Maywood, N J).* 2018; 243(3): 313–322.
28. Kim R, et al. Use of big data in drug development for precision medicine. *Expert Rev Precis Med Drug Dev.* 2016; 1(3): 245-253.
29. Eroglu Z. Combination therapy with BRAF and MEK inhibitors for melanoma: latest evidence and place in therapy *Ther Adv Med Oncol.* 2016; 8(1): 48–56.
30. Serati M, et al. Systematic Review of Combination Drug Therapy for Non-neurogenic Lower Urinary Tract Symptoms *Eur uro.* 2019;75(1): 129–168.
31. Costa F, et al. Big data in biomedicine. *Drug Discovery Today.* 2014;19(4): 433-440.
32. Dash S, et al. Big data in healthcare: management, analysis and future prospects. *J Of Big Da.* 2019; 6(1).
33. Hulsen T, et al. From Big Data to Precision Medicine *Front Med.* 2019; 6:34.
34. Zhu H, et al. Big Data and Artificial Intelligence Modeling for Drug Discovery *Annu Rev Pharmacol Toxicol.* 2020; 60: 573–589.