

Big Data Analytics in urology: the story so far and the road ahead

B. M. Zeeshan Hameed, Aiswarya V. L. S. Dhavileswarapu, Nithesh Naik^{ID}, Hadis Karimi, Padmaraj Hegde, Bhavan Prasad Rai and Bhaskar K. Somani^{ID}

Abstract

Abstract: Artificial intelligence (AI) has a proven record of application in the field of medicine and is used in various urological conditions such as oncology, urolithiasis, paediatric urology, urogynaecology, infertility and reconstruction. Data is the driving force of AI and the past decades have undoubtedly witnessed an upsurge in healthcare data. Urology is a specialty that has always been at the forefront of innovation and research and has rapidly embraced technologies to improve patient outcomes and experience. Advancements made in Big Data Analytics raised the expectations about the future of urology. This review aims to investigate the role of big data and its blend with AI for trends and use in urology. We explore the different sources of big data in urology and explicate their current and future applications. A positive trend has been exhibited by the advent and implementation of AI in urology with data available from several databases. The extensive use of big data for the diagnosis and treatment of urological disorders is still in its early stage and under validation. In future however, big data will no doubt play a major role in the management of urological conditions.

Keywords: artificial intelligence, big data, cancer, data analytics, genome, healthcare, kidney calculi, oncology, urology

Received: 27 January 2021; revised manuscript accepted: 4 February 2021.

Introduction

Big data has garnered a lot of interest among clinicians in the current scenario. There is no denying the fact that almost every sector in the world today is driven by data and the healthcare industry is no exception. Advances in medical technology, electronic medical databases and computational capacity are generating big data in the field of medicine.¹ This massive quantity of data obtained also involves information from devices as small as ingestible sensors, smartphones and watches, along with a variety of electronic health data sets. The electronic health record (EHR) enables big data in the health industry as patient care is routinely documented in EHRs. The data is then fed to large repositories which grow in size and scope, becoming big data resources.² Analysing the information presented in this data may reveal connotations, patterns and trends to progress patient care and reduce costs.³

With the data emerging at an exponential rate, the complexity of dealing with and utilizing it increases. This leads to difficulties in offering personalized treatment plans.⁴ To offer a precise and transversal view of a clinical scenario, artificial intelligence (AI) with machine learning (ML) algorithms and artificial neural networks (ANNs) process was adopted. This soon had a promising wide application and urology is one such area where AI is being widely adopted.⁵

Urology is a specialty that has always been at the forefront of innovation and research where technologies have been rapidly embraced, and this has helped achieve better patient outcomes.⁶ It is one of the most rapidly expanding surgical super specialties and AI paired up with big data plays an important role behind its exponential propulsion. The scientific breakthroughs have certainly helped over the past 20 years, where AI has been

Ther Adv Urol

2021, Vol. 13: 1–13

DOI: 10.1177/
1756287221998134

© The Author(s), 2021.
Article reuse guidelines:
sagepub.com/journals-
permissions

Correspondence to:

Nithesh Naik

Department of Mechanical and Manufacturing Engineering, Faculty of Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

iTRUE (International Training and Research in Uro-Oncology and Endourology) Group
nithesh.naik@manipal.edu

B M Zeeshan Hameed

Department of Urology, Kasturba Medical College Manipal, Manipal Academy of Higher Education, Manipal, India KMC Innovation Centre, Manipal Academy of Higher Education, Manipal, India iTRUE (International Training and Research in Uro-Oncology and Endourology) Group

Aiswarya VLS

Dhavileswarapu
Department of Electronics and Communication, GITAM University, Gandhi Nagar, Rushi Konda, Visakhapatnam, Andhra Pradesh, India

Hadis Karimi

Department of Pharmacy, Manipal College of Pharmaceutical Sciences, Manipal Academy of Higher Education, Manipal, India

Padmaraj Hegde

Department of Urology, Kasturba Medical College Manipal, Manipal Academy of Higher Education, Manipal, India

Bhavan Prasad Rai

iTRUE (International Training and Research in Uro-Oncology and Endourology) Group
Department of Urology, Freeman Hospital, Newcastle, UK

Bhaskar K Somani

Department of Urology, Kasturba Medical College Manipal, Manipal Academy of Higher Education, Manipal, India



iTRUE (International Training and Research in Uro-oncology and Endourology) Group
Department of Urology,
University Hospital
Southampton NHS Trust,
Southampton, UK

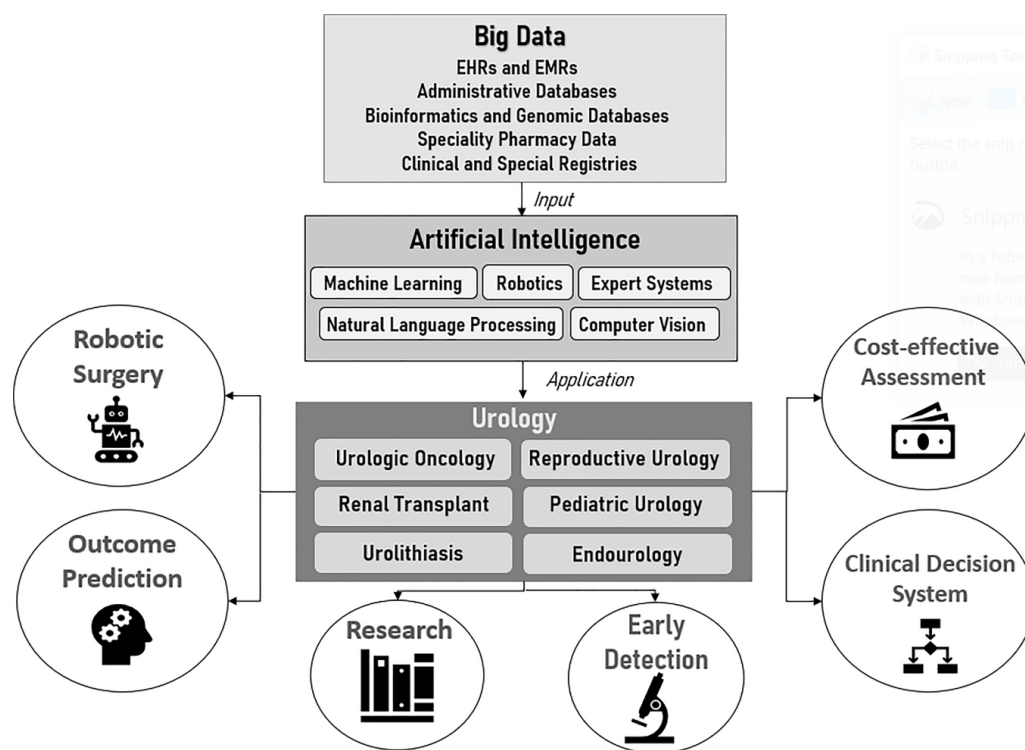


Figure 1. Application of Big Data Analytics and artificial intelligence in healthcare. EHR, electronic health record; EMR, electronic medical record.

extensively applied for the diagnosis,⁷ management⁸ and outcome prediction^{8,9} of urological diseases and conditions (Figure 1).

AI systems are armed with a lot of information to assist in clinical decision making in both predictive and prescriptive analysis. Therefore understanding what exactly big data is and how it is used in these AI applications for urology is of utmost importance. In this review, we explore the major sources of big data used for the advancements in urology and explicate their current and future applications.

Search strategy and article selection

A non-systematic review of all urology related English language literature published in the last decade (2010–2020) was conducted in June 2020 using *MEDLINE*, *Scopus*, *EMBASE* and *Google Scholar*. Our search strategy involved creating a search string based on a combination of keywords. They were: ‘Big Data’, ‘Big Data Analytics’, ‘Urology’, ‘Artificial Intelligence’, ‘AI’, ‘Machine learning’, ‘ML’, ‘ANN’, ‘Convolutional Networks’, ‘Electronic Health Records’, ‘EHR’, ‘EMR’, ‘Bioinformatics’,

‘Genome’, ‘Prostate cancer’, ‘Urinary incontinence’, ‘Kidney stone disease’, ‘Ureteric stones’, ‘Infertility’, ‘Andrology’, ‘Renal cell carcinoma’, ‘Paediatric urology’ and ‘Bladder cancer’. We included original articles published in English.

Inclusion criteria

1. Articles on Big Data Analytics, urology and AI;
2. Full-text original articles on all aspects of diagnosis, treatment and outcomes of urological disorders.

Exclusion criteria

1. Commentaries, reviews and articles with no full text context and book chapters;
2. Animal, laboratory or cadaveric studies.

The literature review was performed as described above. The evaluation of titles and abstracts, screening, and the full article text was conducted for the chosen articles that satisfied the inclusion criteria. Furthermore, the authors manually reviewed the selected articles’ reference lists to screen for any additional work of interest. The

authors resolved the disagreements about eligibility for a consensus decision after discussion.

Big Data Analytics in urology

Digitization of healthcare in recent times led to the generation of large amounts of health data on a day-to-day basis. The data produced is beyond manageable by the traditional software and hardware in terms of storage, processing and analysis, thus rightly being given the name 'big data'.^{10,11} In simple words, big data in healthcare corresponds to the digitally collected patient data amassed from numerous sources including EHRs, medical imaging and genomic sequencing to name a few. The difficulty in harnessing big data is a result of its characteristics – volume (amount of data), variety [type – (structured or unstructured), format – (images, text, video, audio)] and velocity (the increasing rate of data accumulation).² To understand where exactly the data is acquired from and how it contributes to urology in particular, it is essential to discuss the different sources of big data in urology and their respective applications.

Sources and their utilization

EHRs and electronic medical records

EHRs are considered to be the most appropriate form of clinical data available. They comprise the patient's medical history, diagnosis, medications and treatment plans, allergies, imaging data, laboratory reports, test results and clinical outcomes. In short, they are a comprehensive report of a patient's entire health information that can be accessed by authorized users whenever and wherever in the world. Their relevancy compared with any other source of big data in healthcare comes from the fact that they are patient-centred and are created by authorized professionals with the sole purpose of supporting interoperability between health organizations. An ideal EHR system is one that improves aggregation, analysis and communication of patient information.¹¹

Often confused with EHRs, electronic medical records (EMRs) on the other hand are digitized patient charts that are limited to one practice itself. It contains the medical and treatment history of a patient within one practice alone. These are used by the provider for early diagnosis and treatment, unlike EHRs, which are highly used

for decision-making. The main aim of EMR systems is to enhance the quality of care by utilizing its information for various tasks, from scheduling patient appointments to monitoring vital parameters.¹² When choosing a particular EMR system for their practices, the providers must check with the system's features. There are some accomplishments an efficient EHR system is expected to achieve, the most important being privacy for patient data. Figure 2 depicts nine crucial features to look for in the right EHR/EMR.

Traditionally, EMR vendors were fixated upon delivering general-purpose systems that can be used across different specialties. This led to the generation of several gaps within the collected data, thus failing to capture precise data related to a particular disease state. Though such limitations initially hindered the usage of EMRs and EHRs due to lack of important features and inefficient design of the systems, various add-on data analytics platforms were introduced to mitigate these difficulties.^{11,12} Along with enabling patient identification and population management, incorporating data analytics into EMR systems provided visibility into clinical data such as symptom scores and medication utilization.¹² Consequently, workflows could be created targeting the highest-priority patients first and delivering appropriate care to them promptly and more efficiently.

In the present-day scenario, many urology-based EHR systems are available that primarily focus on gathering disease-specific information from the patient. With the existence of urology-specific EHR templates for conditions such as recurrent urinary tract infections, benign prostate disorders, urolithiasis, uro-oncology and many more, extracting relevant information for studies and research has become easier. Focusing on patient-centred outcomes, Tina *et al.*¹³ used an EHR system to detect urinary incontinence following prostatectomy, highlighting how the data captured in EHRs can be used to assess disease treatment. Other similar studies that made use of existing hospital EHR and EMR systems are discussed in Table 1.

The studies shown in Table 1 emphasize how data from EHRs and EMRs, known as big data, contribute to deriving significant insights related to patients and urological diseases. With frequent upgrades in technology, increasing the adoption

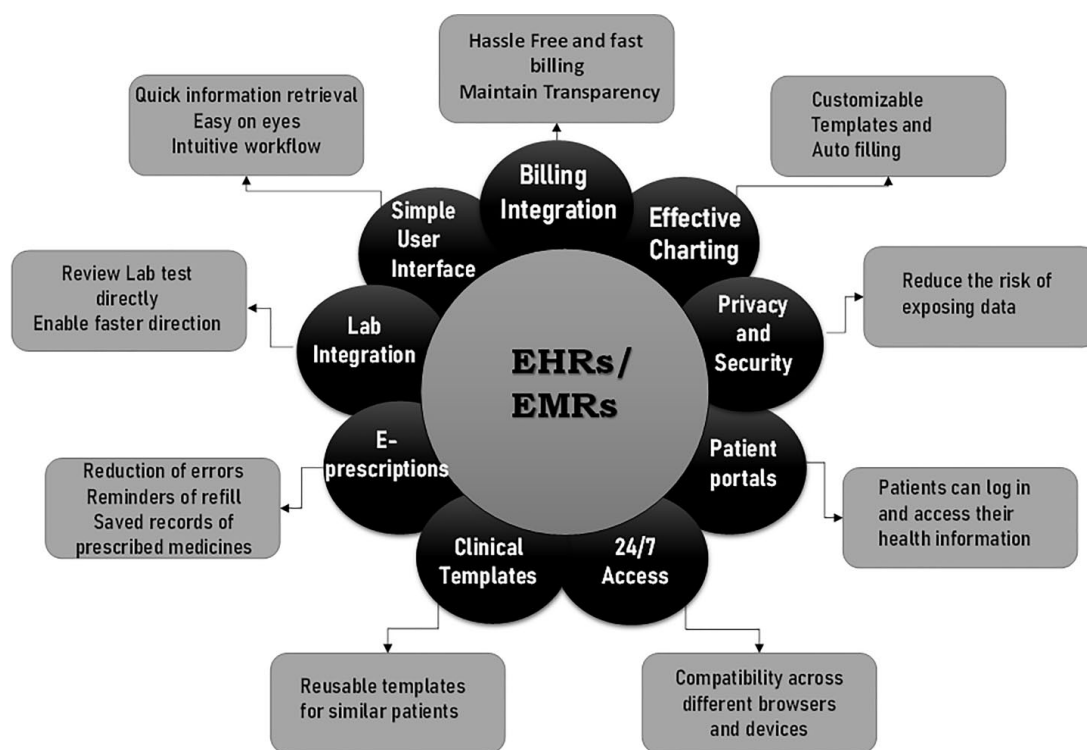


Figure 2. Nine crucial features of EHR/EMR.
EHR, electronic health record; EMR, electronic medical record.

of certified urology EMR and EHR systems in practices enables several advantages to move ahead and remain financially competitive in a healthcare setting.

Administrative data

Administrative data (also known as routinely collected data) is another source of big data in healthcare that is highly employed to inform clinical research.³³ Unlike EMRs or EHRs, administrative data (AD) is primarily collected for reasons other than research (financial aspects of healthcare) and usually consists of enrolment data, hospital in-patient and out-patient data, health insurance claims and pharmacy data.^{33,34} Typically, the data obtained from AD is used to determine and analyse national healthcare utilization trends, access, charges, quality and outcomes.³⁴ Though EMRs offer an advantage over AD in terms of possessing more informed patient details, assessing primary care process quality measures, laboratory test ordering or prescriptions, using it for secondary purposes is not advisable.^{35,36} Therefore, secondary data analysis is most commonly applied to AD.

Secondary data analysis is leveraging the data for research traditionally collected by someone other than the investigator.³⁷ Especially in urological literature, there has been a dramatic increase in utilizing secondary data analysis for clinical research.³⁷ *NIS* (*National Inpatient Sample*) and *KID* (*Kid's Inpatient Database*) derived from samples of the *SID* (*State Inpatient Database*) are some examples of nationally representative discharge data sets that employ secondary data analysis.³⁴ *PHIS* (*Paediatrics Health Information System*) and *National Surgical Quality Improvement Data (NSQID)* are some popular administrative databases used in urology. The latter comprises more than 100 data points and is widely used in urological studies. Various other AD sources and the reviews of urology contingent on them are discussed in Table 1.

Gathered to analyse and deal with the financial burden of diseases, AD has certain limitations that include difficulty in access and use of incomprehensive information on diagnosis and uncertainty regarding its generalizability. Using EMR (clinical) data as a reference standard for AD (financial) could facilitate in providing a

Table 1. Urology studies based on various available databases and registries.

Source	Author	Objective	Disease/subfield	Sample size	Algorithm/model	Outcomes	Results
Urology studies based on EHR/EMR data							
EHR	Tina <i>et al.</i> ¹⁴	Predict factors associated with urinary incontinence following prostatectomy for prostate cancer	Urinary incontinence	3792 Prostate Cancer (PCa) patients who underwent prostatectomy	Natural language processing	1. Significant association between preoperative Urinary Incontinence (UI) and UI is found in the first and second years following surgery 2. Body mass index is found to be a factor associated with UI in the second year	1. First year – odds ratio (OR): 2.30, 95% confidence interval (CI) 1.24–4.28 Second year – OR: 2.24, 95% CI 1.04–4.83 2. OR: 1.11, 95% CI 1.02–1.21
EHR	Andrew <i>et al.</i> ¹⁵	To identify prostate cancer patients that may qualify for active surveillance (AS)	Prostate cancer (uro-oncology)	649 patients who received prostate biopsy	Evidence-based criteria for detecting AS candidates using EHR	Implementation of evidence-based criteria for detection of AS candidates is feasible using EHR data and provides a reasonable basis for delivery system evaluation of practice patterns and for quality improvement	Estimated guideline adherence measured using area under the curve (AUC) was 0.70 (95% CI: 0.66–0.73)
EHR	Ruth <i>et al.</i> ¹⁶	To determine the utility of using routinely collected EHR data for multi-centre analysis of variables predictive of patient no-shows (NSs) to identify areas for future intervention	Paediatric urology	28,715 from Children's Hospital Colorado, Rady Children's Hospital San Diego, and University of Virginia Hospital	Automated electronic data extraction techniques Multivariate logistic regression	1. A total of 2994 NS patients were identified from our sample size, with a mean NS rate of 10.4% 2. Appointment with (a) mid-level provider and (b) huge difference between the dates of scheduling and appointment were found to be significant factors associated with NS appointments	(a) OR: 1.70 95% CI 1.56–1.85 (b) 15–28 days, OR: 1.24, CI 1.09–1.41 29+ days, OR: 1.70, CI 1.53–1.89
EMR	Seneviratne <i>et al.</i> ¹⁷	Identifying cases of metastatic prostate cancer using machine learning on EHRs	Prostate cancer (uro-oncology)	Cohort of 5861 prostate cancer patients	Observational Medical Outcomes Partnership model Random forest model	The developed (A) random forest model outperformed (B) ICD code-based queries in identifying patients with metastatic disease in terms of precision Thus, high-precision classifiers can still assist in identifying cohorts for observational research or clinical trial matching	(A) Precision, 0.90 Recall, 0.40 (B) Precision, 0.33 Recall, 0.54
EMR	Gaylis <i>et al.</i> ¹⁸	To evaluate the potential of an exam-based tumour staging template embedded within an EMR to improve consistency and clarity of clinical tumour staging	Prostate cancer (uro-oncology)	573 biopsies	Implement a staging template	For both men (A) at risk with prostate cancer and (B) with positive biopsies (a) Explicit staging increased (b) Overall staging increased (c) Unconfident staging decreased	A. (a) 16–60% (b) 74–92% (c) 26–8% B. (a) 29–64% (b) 76–92% (c) 24–8%
Urology studies based on administrative databases							
The Healthcare Cost and Utilization Project (HCUP) State Inpatient Database (SID)	Li <i>et al.</i> ¹⁹	To estimate who is at risk of revisiting post prostate brachytherapy	Prostate cancer (uro-oncology)	9042 patients who underwent brachytherapy for PCa	Logistic regression	The factors indicating the higher risk of revisit are: a) Old age b) Recent inpatient admissions (within 3 months before surgery) c) Recent Emergency Department encounters (within 6 months before surgery)	(a) 65–75 years: OR: 1.3, 95% CI 1.06–1.60, $p = 0.001$ >75 years: OR: 1.5, 95% CI 1.18–1.97, $p = 0.001$ (b) OR: 2.68, 95% CI 1.8–4.0, $p < 0.001$ (c) OR: 1.63, 95% CI 1.4–1.89, $p < 0.001$

(Continued)

Table 1. (Continued)

Source	Author	Objective	Disease/subfield	Sample size	Algorithm/model	Outcomes	Results
<i>Clinformatics® Data Mart (CDM)</i> (consists of administrative health claims data)	Ward <i>et al.</i> ²⁰	To investigate the recent epidemiology of the paediatric urinary stone disease (USD) in the USA	Paediatric urology	12,739,125 children (0–18 years) who were identified to suffer from USD	N/A	The study offers one of the most extensive assessments of paediatric USD to date, exhibiting changing rates and shifting treatment patterns, just as contrasts by age, sex, race/ethnicity and geographic zones	USD rate for 2005–2016 period was found to be 59.5 cases per 100,000 person-years
<i>HCUP: SID SASD SEDD</i>	Shah <i>et al.</i> ²¹	To quantify the risk of opioid dependence (OD) or overdose among patients undergoing urologic surgery and to identify risk factors of OD or overdose	Urologic surgery	675,525 patients who underwent urological surgery	Multivariable logistic regression	1 out of 1111 patients who underwent urological procedures is found to be a victim of postoperative OD The factors associated with OD are: (a) Younger age (b) Inpatient procedures (c) Increasing hospitalization duration (d) Baseline depression (e) Chronic obstructive pulmonary disease (f) Has non-private insurance	(a) 51 versus 62 median age (b) 81.0% versus 42.4% (c) 3 versus 0 days median (d) 14.4% versus 3.4% (e) 20.3% versus 8.9% (f) 69.6% versus 66%
<i>National Inpatient Sample (NIS) Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS)</i> data	Shirk <i>et al.</i> ²²	To analyse the dependency of urologic surgery outcomes to patient experience	Urologic oncology and surgery	A cohort of 46,988 who were admitted for cancer-directed prostatectomy, nephrectomy and cystectomy	N/A	Patient experience is found to be a factor that is best suitable in quality analysis rather than for improving surgical outcomes In fact, length of hospitalization and nursing-sensitive complications are highly dependent on the performance of the hospital Higher performance, reduced stay and complications	Hospitalization OR: 0.77, 95% CI 0.64–0.92 Nursing-sensitive complications OR: 0.85, 95% CI 0.72–0.99
<i>NIS</i>	Qin <i>et al.</i> ²³	To evaluate the perioperative outcomes between robot-assisted radical prostatectomy (RARP) and open radical prostatectomy (ORP)	Prostate cancer (urologic oncology)	78,440 patients who underwent RARP surgery	Logistic regression	(a) RARP outperformed ORP in certain aspects (b) It is also found that there is a possible diminishing propensity of complexity incidence rates for the same aspects as above in RARP	(a) [RARP versus ORP] Annual per cent change (APC) in RARP (i) Blood transfusion (a) 1.96% versus 9.40% (b) APC = -9.81 (ii) Intraoperative complication (a) 0.73% versus 1.25% (b) APC = -12.84 (iii) Overall postoperative complications (a) 8.87% versus 11.97% (b) APC = -14.09 (iv) Prolonged length of stay (a) 13.39% versus 36.70%
<i>NIS and Nationwide German Hospital Billing Database (NGHBD)</i>	Groeben <i>et al.</i> ²⁴	To estimate and compare trends of urinary diversion (UD) for patients receiving radical cystectomy (RC) for bladder cancer in USA and Germany, and further evaluate decisive predictors for the choice of UD	Bladder cancer (urologic oncology)	17,711 RC cases from NIS and 60,447 RC cases from NGHBD	Descriptive analysis Logistic regression (evaluate possible factors related to incontinent UD) Linear regression (trend detection)	(a) Increasing age of patients with probably higher comorbidity lately expanded the utilization of incontinent UD in Germany, while continent UD gives an impression of being underused in the USA (b) In-hospital mortality was lower in the USA compared with Germany	(a) Incontinent UD USA: stable at 93% Germany: increase from 63.2% to 70.8% (b) 1.9% in USA versus 4.6% in Germany, $p=0.001$ Hospital stay: 10.7 days in the USA versus 25.1 days in Germany, $p=0.001$

(Continued)

Table 1. (Continued)

Source	Author	Objective	Disease/subfield	Sample size	Algorithm/model	Outcomes	Results
Paediatrics Health Information System (PHIS)	Suson et al. ²⁵	To compare the contemporary presentation and results of all-cause nephrectomy performed by paediatric urologists (PUs) and paediatric general specialists (PGSs)	Urologic surgery and general surgery	6520 nephrectomy cases	N/A	(a) Compared with PGSs more nephrectomies are performed by PUs (b) Patients operated by PUs were observed with: (i) Lower severity level and at (ii) Lower risk of mortality than those operated by PGSs (c) Those treated with nephrectomy for benign disease by PU compared with PGS had: (i) Shorter length of stay (ii) Lesser medical and (iii) Post-procedure complications	(a) PUs performed 61% more nephrectomies than PGSs (b) Malignant: (i) $p < 0.013$ (ii) $p < 0.008$ Benign: $p < 0.001$ (c) (i) $p < 0.001$ (ii) $p = 0.009$ (iii) $p = 0.001$
National Surgical Quality Improvement Program (NSQIP)	Vij et al. ²⁶	To implement a simple operating room bundle for reducing surgical site infection (SSI) rate post urologic surgery	Urologic surgery	510 urologic cases such as cystectomy, nephrectomy and prostatectomy	N/A	Introducing a straightforward, quick, cheap and effectively reproducible bundle resulted in a diminished rate of SSI post major urologic medical procedures. Utilization of this bundle could be beneficial considering how lower SSI rates are critical for patients and physicians of this era of outcome-based reimbursements	Before the introduction of bundle: SSI=3.57% After the implementation of bundle: SSI=1.37% $p = 0.23$
Urology studies based on genome database and clinical registries							
National Institutes of Health, Online Mendelian Inheritance in Man and OrphaNet	Cariati et al. ²⁷	To highlight the importance of molecular testing in reproductive medicine	Male and female infertility (reproductive urology)	285 patients severe oligospermia and non-obstructive azoospermia	N/A	Genetic testing could help in analysing male and female infertility and thus should be highly encouraged by the reproductive specialists	NA
Genome-wide association studies (GWAS) summary statistics, UK Biobank and BioMe Biobank	Paranjpe et al. ²⁸	To analyse the association of urinary tract stones (UTSs) with polygenic risk scores (PRSs)	UTSs	28,877	Logistic regression	Genome wide PRSs are related to UTSs and thus can be utilized for early stratification in the absence of other clinical risk factors OR is found to have linear relationship with increasing standard deviation in PRSs	(a) Overall OR: 1.2, 95% CI 1.13–1.26 (b) Top decile compared with lowest decile group OR: 2.6, 95% CI 1.9–3.6 (c) Low risk group OR: 1.3, 95% CI 1.12–1.58 (d) High risk group OR: 1.2, 95% CI 1.1–1.2
Genetic data directly taken from Saarland University Hospital and Institute of Pathology of the University Hospital in Erlangen	Grimm et al. ²⁹	Implement an algorithm to determine metastatic risk stratification of clear cell renal cell carcinoma utilizing genetic information	Renal cell carcinoma in children	200	Logistic regression and multivariate Cox proportional analysis	The authors were able to design a genetic risk score model that gave better accuracy than Leibovich score The model delivered higher accuracy in identifying patients with the malignancy	The accuracy generated by the designed model is 87% with 86% specificity, 88% sensitivity
Surveillance, Epidemiology, and End Results (SEER)	Eminaga et al. ³⁰	To build an artificial intelligence (AI)-based data-driven solution for risk adapted follow-up scheduling that could highly benefit individuals with urologic cancers	Different urologic cancers	2,006,052	Recurrent neural networks Random forest XGBoost Linear Support Vector Machine	AI-based solution is found to be a feasible tool that can be utilized for effective management of urologic cancer patients' surveillance	The overall concordance index score delivered by the model is 0.80

(Continued)

Table 1. (Continued)

Source	Author	Objective	Disease/subfield	Sample size	Algorithm/model	Outcomes	Results
<i>Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE)</i>	Jeong et al. ³¹	To assess and compare surgical outcomes and changes in urinary and sexual quality of life (QOL) over time in patients who underwent radical prostatectomy (RP) in ordinary (ORP) and with robot assistance (RARP).	Prostate Cancer	1892	Repeated measure mixed models Logistic regression Cox regression	Difference in surgical approach certainly did not change the recovery time post the procedure Most of the patients did experience changes in their urinary and sexual QOL after robot assisted RP similar to that in ordinary RP	(a) Urinary incontinence: ORP Mean \pm standard deviation [M#SD]=69 \pm 26 <i>versus</i> RARP M#SD = 62 \pm 27 (b) Bothing only for a year after RP ORP M#SD = 75 \pm 29 <i>versus</i> RARP M#SD = 68 \pm 28
<i>National Trauma Data Bank (NTDB)</i>	Johnsen et al. ³²	To examine whether placing suprapubic tube (ST) in the patients undergoing internal fixation (IF) for treating their urethral trauma (UT) will lead to infections in the index of hospital stay or not	Urethral injury	969	Poisson regression analysis Univariate and multivariate analysis	The study found there is no risk of infectious complications associated with placing ST for UT patients undergoing IF. In fact, injury severity score (ISS) and smoking are found to be the factors affecting infection issues	(a) ISS: OR:4.0, 95% CI 1.25–12.77 (b) Smoking: OR:2.45, 95% CI 1.11–5.43

comprehensive picture of patient health information that can be utilized to assess outcomes more accurately.³⁸ Similarly, having a clear goal for the study, choosing an appropriate dataset and avoiding ill-fitted statistical analysis could resolve the issues when applying secondary data analysis on data sets.³⁷

Bioinformatics and genomic databases

The dawn of the genomic medicine era triggered unforeseen perception towards genetic variations that drives tumour development and progression.³⁹ The emergence of modern bioinformatics in biomedical research opened up tremendous opportunities to derive powerful insights from the clinically constructed genetic databases.

Today an individual’s entire genome sequence is shared securely over the web. Databases known as genome browsers offer a way of sharing genome information in an accessible format after it is sequenced, assembled and annotated.⁴⁰ Some examples of genome browsers include Ensembl, a joint project between European Bioinformatics Institute (EBI), part of the European Molecular Biology Laboratory (EMBL) and the Wellcome Trust Sanger Institute (in UK), UCSC (genome browser-based from University of California Santa Cruz) and NCBI (National Centre for Biotechnology Information).⁴⁰ Some of the prominent genome databases that make up big data in healthcare dealing with biological information are shown in Table 2.

The genome is a complete set of information in an organism’s DNA.⁴¹ Though the basic concepts involved in discussing genome medicine such as DNA, microRNA, biomarkers and others are challenging to comprehend, understanding them might lead to perceiving various diseases.⁴² This highly aids providing optimum care to the patients by identifying individual risk factors and recommending strategies to counter them in short, personalized treatment. Major advances in genome medicine aim to deliver precision medicine, gene therapy and genetic therapy and contribute to the field of ‘omics’.^{41,42}

Identification of genetic alterations that progress malignant diseases such as prostate cancer facilitates the possibility of personalized medicine. In urology, though genome data was primarily focused on cancer therapy, in recent years there has been a significant influence on non-cancerous diseases such as erectile dysfunction (ED) and

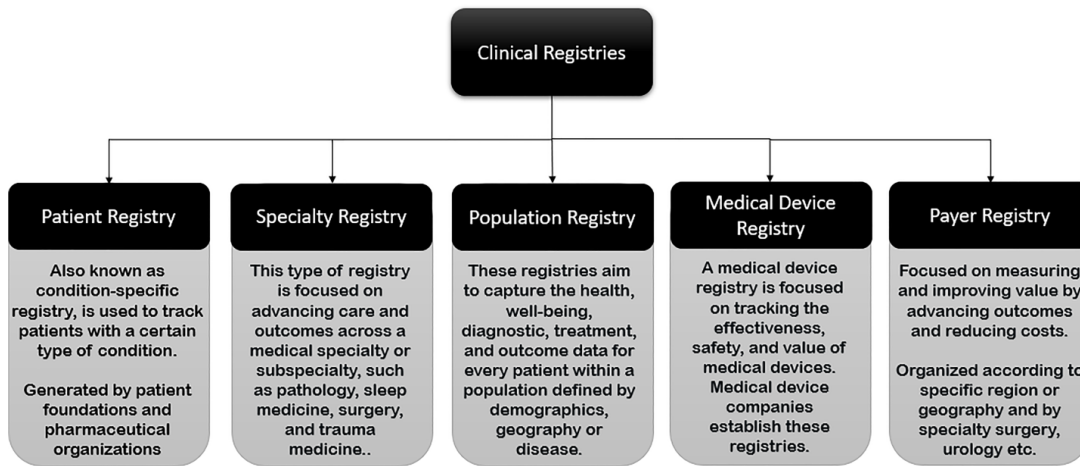


Figure 3. Classification of clinical registries used as healthcare data.

Table 2. Prominent genome browsers available online.

Serial Number	Name	Year of launch	URL	Type of data/resource	Target user
1	NCBI Genome	1988	https://www.ncbi.nlm.nih.gov/genome	Genome sequences, maps, chromosomes, assemblies and annotations	Geneticist
2	Ensembl Genome	1999	http://ensemblgenomes.org/	Variant knowledge base, meta resource	Geneticist, molecular biologists and other researchers
3	GMOD Project	Early 2000s	http://gmod.org/wiki/Main_Page	A deep catalogue of human genetic variations	Geneticists and biologists
4	H-InvDB	2004	http://www.h-invitational.jp/	Human genes and transcripts	Geneticist
5	GWASCentral	2001	https://www.gwascentral.org/	Summary level findings from genetic association studies. Primary focus on single-nucleotide polymorphisms	Genetic counsellors, geneticists
6	dbVar	2012	https://www.ncbi.nlm.nih.gov/dbvar	Genotype and Phenotype	Geneticist
7	UCSC	2000	https://genome.ucsc.edu/	Genome sequence data from a variety of species	Geneticist
8	ENCODE	2003	https://www.encodeproject.org/	All functional elements of human genome	Geneticist
9	IGSR (1000 Genomes)	2008	http://www.internationalgenome.org/	Human genetic variations	Geneticist
10	GenBank	1982	https://www.ncbi.nlm.nih.gov/genbank/	Nucleotide and protein sequences	Genetic counsellors and geneticists

bladder dysfunctions as well. One such case is the study by Patel *et al.* based on identifying causes of ED through genome data.⁴³ Studies shown in Table 1 give an idea of the range of urologic diseases utilizing bioinformatics and genome databases for advancements in disease identification and treatment.

The global availability combined with ease of access to DNA-sequencing data has bestowed upon genetics research an unparalleled potential required to understand diseases and their complex traits.⁴² Inappropriate use of genomic data poses particular risks since it can be used to identify an individual.⁴⁴ Provided such risks can be avoided, or at the least be reduced, large-scale sharing of genome information could help in extending biomedical research and help tackle and potentially help with certain diseases.

Specialty pharmacy data

The main aim of Specialty Pharmacies (SPs) is to provide expert clinical care to people suffering from serious illnesses such as cancer. They are equipped to handle complicated conditions and have access to advanced medications compared with traditional pharmacies. The high-touch services delivered by high-cost and highly complex specialty pharmaceuticals create data (clinical and financial) opportunities that hold an exceptional value amongst their stakeholders. SPs need to collect and aggregate data for their efficient patient management and overall success. The patient data stored at SPs is gathered by direct interaction with a patient through utilization reviews, patient counselling and follow-up care.⁴⁵ This factor makes data from SPs highly valuable for pharmaceutical industries who use it to enhance their drug's efficacy.^{44,45} Strengthening the therapeutic value of the drug not only increases the drug efficacy but also ensures a better patient experience and improves the health of the population.⁴⁵

Urologic oncology is one sub-specialty of urology that SPs highly contribute to, in terms of therapy. For effective treatment of patients with conditions such as prostate cancer, bladder cancer, kidney cancer and other urologic diseases, specialty pharmaceuticals are prescribed by urologists.⁴⁶ There is also a provision of providing additional SP treatment options in the future. SPs' impact on urology is expected to continuously grow as the data generated by SPs continues to benefit

various life-threatening diseases both in urological diseases and other fields of medicine.^{45,46}

Clinical or condition-specific registries

Condition-specific registries are a type of clinical registry with examples such as population registry, specialty registry, medical device registry and payer registry. Each registry typically focuses on collecting information based on a particular aspect.^{46,47} The medical device registry gathers information fit to answer questions concerning the effectiveness, value and safety of medical devices. Similarly, specialty registries are a type of clinical registry that possesses information similar to SPs. While SPs focus on advancing care for a patient of complex diseases, specialty registries concentrate on doing the same with a medical specialty or sub-specialty (such as surgery or pathology).^{37,47} The classification of clinical registries used as healthcare data is shown in Figure 3.

Condition-based registries are large data sets produced from clinical data of patients with a specific type of disease or disorder.³⁷ Unlike administrative data or claims data, condition-specific registries are generated to study and analyse a particular disease condition. Apart from being the primary source of study, these are also often used by urologic investigators for secondary data analysis.³⁷ Some examples of such registries that are mostly used for secondary data analysis are the *SEER (Surveillance, Epidemiology, and End Results)*, *CaPSURE (Cancer of the Prostate Strategic Urologic Research Endeavor)*, and *NTDB (The National Trauma Data Bank)* data sets.³⁷ A few urology studies based on these clinical registries are illustrated in Table 1.

Though registries provide a solution for some issues when using AD, their core limitation of cost restricts their scalability. While both automated and manual (by paid registrars) data abstraction costs a lot, the former is susceptible to inaccuracies as well.⁴⁸ Nonetheless, clinical registries though are recent developments and are likely to play crucial roles in quality improvement and yield studies that will hold a large share of urologic literature given their advantages over AD.⁴⁸

Discussion

These five sources mentioned above hold up for a significant part as big data sources in the healthcare industry, especially in urology. We have

discussed the critical applications of each source for urology and briefly corroborated with the studies listed in their respective tables.^{14–19,26–32} The impact of Big Data Analytics and secondary data analysis on the collected data is evident. Both processes result in discovering associations and hidden patterns in the collected data to prevent epidemics, cure diseases and improve patient quality of life.^{5–10}

Though the real-life implementation of AI remains limited, it has the potential to change the way urology is and will be practised. It enables faster diagnosis and reduction of unnecessary costs in the medical field. Furthermore, AI models are extensively used to enhance treatment efficiency by enabling faster diagnosis, predictive analysis and precision medicine.⁹ The application of novel AI technology in urology has been regarded as a promising step towards improving diagnostic capability and prediction of disease recurrences.⁴⁹ By using highly predictive and accurate AI algorithms, improved diagnoses of male infertility, urinary tract infections and paediatric malformations are possible.⁵⁰ Advancements in technology with the aid of virtual or augmented reality brings in greater potential of AI-assisted surgeries and improves patient care.

While AI is hailed for all these accomplishments, it would not have reached that status without big data. AI and big data are equally important and responsible for the advancements made in urology. Therefore, to offer a complete and clear perspective on the future beheld for urology, this review discusses the prominent big data sources in urology in detail.^{10,11}

Genome data and SP data can truly deliver groundbreaking results in uro-oncology. While the former can help understand the reason behind the disease, the latter enables a chance to deliver improved medication and patient care for advanced conditions. Both of these sources have the utmost significance in providing precision medicine.¹² Along with assisting quality enhancement, condition-specific registries provide extremely relevant clinical data valuable for urological research. EHRs and AD together can provide a broader view to deal with many aspects of the healthcare industry. Compared with traditional statistical models, AI models are considered superior by the majority of surveyed studies. As the construction and management of big data resources develop along with

much more reliable and efficient AI techniques, we believe that there truly will be a transformation in the way urological diseases are dealt with in terms of diagnosis and treatment.^{15–18} With the onslaught of the COVID pandemic, big data is also being used to tackle it and to prioritize mass vaccination programmes.^{51,52}

While data from the Internet of Things (IoT) devices are also considered a major contribution to healthcare data, IoT is still in very early stages, especially in urology. For the sake of brevity, survey and research data, which plays a less significant role compared with other sources, is not discussed in this review. We did not carry out a ‘risk of bias’ assessment in our study, which should also be done in future studies.

Conclusion

The use of Big Data Analytics in urology has seen a quantum jump over the last decade. The emergence of AI and its application in urology using the data available from various databases is showing a promising trend. The generalized utilization of big data for the diagnosis of several urological conditions and their treatment is still in the incipient stage and under validation. However, in the future big data is no doubt going to take a paramount role in the treatment of various urological conditions.

Author Contributions

Study conception and design: Bhavan Prasad Rai, Bhaskar K. Somani, Nithesh Naik

Acquisition of data: Aiswarya V. L. S. Dhavileswarapu, Nithesh Naik, Hadis Karimi

Analysis and interpretation of data: Nithesh Naik, B. M. Zeeshan Hameed, Padmaraj Hegde

Drafting of manuscript: B. M. Zeeshan Hameed, Aiswarya V. L. S. Dhavileswarapu, Hadis Karimi

Critical revision: Padmaraj Hegde, Bhavan Prasad Rai, Bhaskar K. Somani.

Conflict of interest statement

The authors declare that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ORCID iDs

Nithesh Naik  <https://orcid.org/0000-0003-0356-7697>

Bhaskar K Somani  <https://orcid.org/0000-0002-6248-6478>

References

1. Beam A and Kohane I. Big data and machine learning in health care. *JAMA* 2018; 319: 1317.
2. Ghani KR, Zheng K, Wei JT, *et al.* Harnessing big data for health care and research: are urologists ready? *Eur Urol* 2014; 66: 975–977.
3. Murdoch TB and Detsky AS. The inevitable application of big data to health care. *JAMA* 2013; 309: 1351–1352.
4. Chen J, Remulla D, Nguyen JH, *et al.* Current status of artificial intelligence applications in urology and their potential to influence clinical practice. *BJU Int* 2019; 124: 567–577.
5. Checcucci E, Autorino R, Cacciamani GE, *et al.* Artificial intelligence and neural networks in urology: current clinical applications. *Minerva Urol Nefrol* 2019; 72: 49–57.
6. Venkatramani V. Urovision 2020: the future of urology. *Indian J Urol* 2015; 31: 150.
7. Kanagasingam Y, Xiao D, Vignarajan J, *et al.* Evaluation of artificial intelligence-based grading of diabetic retinopathy in primary care. *JAMA Netw Open* 2018; 1: e182665.
8. Drouin SJ, Yates DR, Hupertan V, *et al.* A systematic review of the tools available for predicting survival and managing patients with urothelial carcinomas of the bladder and of the upper tract in a curative setting. *World J Urol* 2013; 31: 109–116.
9. Hung AJ, Chen J and Gill IS. Automated performance metrics and machine learning algorithms to measure surgeon performance and anticipate clinical outcomes in robotic surgery. *JAMA Surg* 2018; 153: 770–771.
10. Cahan EM, Hernandez-Boussard T, Thadaney-Israni S, *et al.* Putting the data before the algorithm in big data addressing personalized healthcare. *NPJ Digit Med* 2019; 2: 1–6.
11. Raghupathi W and Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014; 2: 3.
12. Azzolina JJ. Data analytics in the large urology practice: patient identification, population management, and protocol adherence. *Rev Urol* 2017; 19: 46.
13. Hernandez-Boussard T, Tamang S, Blayney D, *et al.* New paradigms for patient-centered outcomes research in electronic medical records: an example of detecting urinary incontinence following prostatectomy. *EGEMS (Wash DC)* 2016; 4: 1231.
14. Li K, Banerjee I, Magnani CJ, *et al.* Clinical documentation to predict factors associated with urinary incontinence following prostatectomy for prostate cancer. *Res Rep Urol* 2020; 12: 7–14.
15. Knighton AJ, Belnap T, Brunisholz K, *et al.* Using electronic health record data to identify prostate cancer patients that may qualify for active surveillance. *EGEMS (Wash DC)* 2016; 4: 1220.
16. Bush R, Vemulakonda V, Corbett S, *et al.* Can we predict a national profile of non-attendance pediatric urology patients: a multi-institutional electronic health record study. *Inform Prim Care* 2014; 21: 132.
17. Goldenberg SL, Nir G and Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol* 2019; 16: 391–403.
18. Gaylis F, Nasserri R, Swift S, *et al.* Leveraging the electronic medical record improves prostate cancer clinical staging in a community urology practice. *Urol Pract* 2021; 8: 47–52.
19. Li B, Kirshenbaum EJ, Blackwell RH, *et al.* Thirty-day hospital revisits after prostate brachytherapy: who is at risk? *Prostate Int* 2019; 7: 68–72.
20. Ward JB, Feinstein L, Pierce C, *et al.* Pediatric urinary stone disease in the United States: the urologic diseases in America project. *Urology* 2019; 129: 180–187.
21. Shah AS, Blackwell RH, Kuo PC, *et al.* Rates and risk factors for opioid dependence and overdose after urological surgery. *J Urol* 2017; 198: 1130–1136.
22. Shirk JD, Tan HJ, Hu JC, *et al.* Patient experience and quality of urologic cancer surgery in US hospitals. *Cancer* 2016; 122: 2571–2578.
23. Qin Y, Han H, Xue Y, *et al.* Comparison and trend of perioperative outcomes between robot-assisted radical prostatectomy and open radical prostatectomy: nationwide inpatient sample 2009–2014. *Int Braz J Urol* 2020; 46: 754–771.
24. Groeben C, Koch R, Baunacke M, *et al.* Robots drive the German radical prostatectomy market: a total population analysis from 2006 to 2013. *Prostate Cancer Prostatic Dis* 2016; 19: 412–416.
25. Suson KD, Wolfe-Christensen C, Elder JS, *et al.* National practice patterns and outcomes of pediatric nephrectomy: comparison between

- urology and general surgery. *J Urol* 2015; 193: 1737–1742.
26. Vij SC, Kartha G, Krishnamurthi V, *et al.* Simple operating room bundle reduces superficial surgical site infections after major urologic surgery. *Urology* 2018; 112: 66–68.
 27. Cariati F, D'Argenio V and Tomaiuolo R. The evolving role of genetic tests in reproductive medicine. *J Transl Med* 2019; 17: 267.
 28. Paranjpe I, Tsao N, Judy R, *et al.* Derivation and validation of genome wide polygenic score for urinary tract stone diagnosis. *Kidney Int* 2020; 98: 1323–1330.
 29. Grimm J, Zeuschner P, Janssen M, *et al.* Metastatic risk stratification of clear cell renal cell carcinoma patients based on genomic aberrations. *Genes Chromosomes Cancer* 2019; 58: 612–618.
 30. Eminaga O, Breil B, Semjonow A, *et al.* Artificial intelligence-based personalized and risk-adapted surveillance management for urologic cancer: a SEER-based study. 2020. DOI: 10.21203/rs.3.rs-52678/v1.
 31. Jeong CW, Cowan JE, Broering JM, *et al.* Robust health utility assessment among long-term survivors of prostate cancer: results from the cancer of the prostate strategic urologic research endeavor registry. *Eur Urol* 2019; 76: 743–751.
 32. Johnsen NV, Vanni AJ and Voelzke BB. Risk of infectious complications in pelvic fracture urethral injury patients managed with internal fixation and suprapubic catheter placement. *J Trauma Acute Care Surg* 2018; 85: 536–540.
 33. Cadarette SM and Wong L. An introduction to health care administrative data. *Can J Hosp Pharm* 2015; 68: 232.
 34. Steiner C, Elixhauser A and Schnaier J. The healthcare cost and utilization project: an overview. *Eff Clin Pract* 2002; 5: 143–151.
 35. Roth CP, Lim YW, Pevnick JM, *et al.* The challenge of measuring quality of care from the electronic health record. *Am J Med Qual* 2009; 24: 385–394.
 36. Rea S, Pathak J, Savova G, *et al.* Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn project. *J Biomed Inform* 2012; 45: 763–771.
 37. Schlomer BJ and Copp HL. Secondary data analysis of large data sets in urology: successes and errors to avoid. *J Urol* 2014; 191: 587–596.
 38. Tu K, Mitiku TF, Ivers NM, *et al.* Evaluation of Electronic Medical Record Administrative data Linked Database (EMRALD). *Am J Manag Care* 2014; 20: e15–e21.
 39. Bernard B and Flaig TW. POINT—Prostate cancer genomic analysis: routine or research only? *Prostate* 2018; 32: 607–609.
 40. How are sequenced genomes stored and shared? Facts <https://www.yourgenome.org/facts/how-are-sequenced-genomes-stored-and-shared> (2016, accessed January 20, 2021).
 41. Roth SC. What is genomic medicine? *J Med Libr Assoc* 2019; 107: 442–448.
 42. Eccles MR, Bailey RR, Abbott GD, *et al.* Unraveling the genetics of vesicoureteric reflux: a common familial disorder. *Hum Mol Genet* 1996; 5: 1425–1429.
 43. Patel DP, Pastuszak AW and Hotaling JM. A review of genome wide association studies for erectile dysfunction. *Curr Sex Health Rep* 2019; 11: 342–347.
 44. Balaji D and Terry SF. Benefits and risks of sharing genomic information. *Genet Test Mol Biomarkers* 2015; 19: 648–649.
 45. Steagall A. How Specialty Pharmacy Data Can Boost Your Drug's Success. McKesson <https://www.mckesson.com/blog/how-specialty-pharmacy-data-can-boost-your-drugs-success/> (accessed 20 January 2021).
 46. Pereira-Azevedo N, Carrasquinho E, De Oliveira EC, *et al.* mHealth in urology: a review of experts' involvement in app development. *PLoS One* 2015; 10: e0125547.
 47. Moore B. What is a clinical data registry and why is it important? //ArborMetrix <https://www.arbormetrix.com/blog/clinical-data-registry-basics> (2020, accessed 4 December 2020).
 48. Tyson MD and Barocas DA. Improving quality through clinical registries in urology. *Curr Opin Urol* 2017; 27: 375.
 49. Shah M, Naik N, Somani BK, *et al.* Artificial Intelligence (AI) in urology – current use and future directions: an iTRUE study. *Turk J Urol* 2020; 46(Suppl. 1): S27–S39.
 50. Checcucci E, De Cillis S, Granato S, *et al.* Applications of neural networks in urology: a systematic review. *Curr Opin Urol* 2020; 30: 788–807.
 51. Ho HC, Hughes T, Bozlu M, *et al.* What do urologists need to know: diagnosis, treatment, and follow-up during COVID-19 pandemic. *Turk J Urol* 2020; 46: 169–177.
 52. Kent J. Intersection of big data analytics, COVID-19 top focus of 2020, 0 <https://healthitanalytics.com/news/intersection-of-big-data-analytics-covid-19-top-focus-of-2020> (accessed 26 January 2021).