Nursing-Relevant Patient Outcomes and Clinical Processes in Data Science Literature: 2019 Year in Review

Mary Anne Schultz, BSN, MSN, MBA, PhD,
California State University

Rachel Lane Walden, MLIS,
Vanderbilt University, Annette and Irwin Eskind Family Biomedical Library

Kenrick Cato, RN, PhD, CPHIMS, FAAN [Assistant Professor],
Columbia University School of Nursing, Department of Emergency Medicine

Cynthia Peltier Coviak, PhD, RN, FNAP [Professor Emeritus],
Grand Valley State University

Christopher Cruz, MSHI, RN-BC, CPHIMS [Team Leader],
Global Health Technology & Informatics, Chevron, San Ramon, CA

Fabio D'Agostino, PhD, MSN, RN [Assistant Professor],
Saint Camillus International University of Health Sciences, Rome, Italy

Brian J. Douthit, MSN, RN-BC,
Duke University School of Nursing

Thompson Forbes, PhD, RN [Assistant Professor],
East Carolina University College of Nursing

Grace Gao, PhD, DNP, RN-BC,
St Catherine University Department of Nursing

Mikyoung Angela Lee, PhD, RN [Associate Director of PhD Program and Associate Professor],
Texas Woman's University College of Nursing

Deborah Lekan, PhD, RN-BC,
University of North Carolina at Greensboro School of Nursing

Ann Wieben, MS, BSN, RN-BC,
University of Wisconsin School of Nursing

Alvin D. Jeffery, PhD, RN-BC, CCRN-K, FNP-BC [Assistant Professor]
Vanderbilt University School of Nursing; Nurse Scientist, Tennessee Valley Healthcare System,
U.S. Department of Veterans Affairs

Abstract

Corresponding Author: Alvin D. Jeffery, 461 21st Ave South, Nashville, TN 37240; 615-322-4400; alvinjeffery@gmail.com.

Data science continues to be recognized and used within healthcare due to the increased availability of large data sets and advanced analytics. It can be challenging for nurse leaders to remain apprised of this rapidly changing landscape. In this paper, we describe our findings from a scoping literature review of papers published in 2019 that use data science to explore, explain, and/or predict 15 phenomena of interest to nurses. Fourteen of the 15 phenomena were associated with at least one paper published in 2019. We identified the use of many contemporary data science methods (e.g., natural language processing, neural networks) for many of the outcomes. We found many studies exploring *Readmissions* and *Pressure Injuries*. The topics of *Artificial Intelligence/Machine Learning Acceptance, Burnout, Patient Safety,* and *Unit Culture* were poorly represented. We hope the studies described in this paper help readers: (a) understand the breadth and depth of data science's ability to improve clinical processes and patient outcomes that are relevant to nurses and (b) identify gaps in the literature that are in need of exploration.

Keywords

Data Analytics; Artificial Intelligence; Nursing Research; Outcome and Process Assessment

INTRODUCTION

The phrase *data science,* along with related phrases like *artificial intelligence, predictive analytics,* and *machine learning*, are increasingly common not only in lay news and media but also in biomedical and nursing literature. One hopes the increasing use of large data sets and advanced analytics is associated with improvements in clinical care delivery and patient outcomes. Unfortunately, the ever-expanding corpus of publications and the plethora of potential clinical applications can leave many nurse leaders struggling to remain apprised of the most current evidence. In this paper, we describe a representative selection of papers published in 2019 that use data science to explore, explain, and/or predict phenomena of interest to nurses.

This project was based on interest from members of the Data Science Workgroup of the Nursing Knowledge: Big Data Science Conference [1] hosted annually by the University of Minnesota School of Nursing. Using a concept analysis paper [2] and group consensus, we identified 15 nursing-relevant patient outcomes and clinical process measures where data science techniques could be helpful. The outcomes selected for review comprise (in alphabetical order): Artificial Intelligence/Machine Learning Acceptance, Burnout, Emergency Department Visits, Falls, Healthcare-Acquired Infections, Healthcare Utilization and Costs, Hospitalization, In-Hospital Mortality, Length of Stay, Pain, Patient Safety, Pressure Injuries, Readmissions, Staffing/Scheduling/Workload, and Unit Culture.

METHODS

A scoping literature review was conducted using PubMed and CINAHL databases in December of 2019 for English language studies published during the past year. The species filter was also used to restrict to human studies. There was one main search strategy which used a combination of keywords and subject headings to find studies discussing the use of data science. The following terms were used to create that strategy: data science,

data analytics, artificial intelligence, machine learning, risk assessment, decision support techniques, clinical prediction rule, natural language processing (NLP), computer-assisted image processing, along with analytic, forecast, prediction, risk, and statistical models. This main strategy was combined with an outcome specific strategy for all 15 outcomes (see Table, Supplemental Digital Content 1, which presents full search strategies). Each outcome was reviewed by an individual author who is an expert in the outcome reviewed. Abstract and full-text screening were done using the Raayan [3] web application. Inclusion/exclusion criteria were developed via group consensus with the intention of providing a representative sample of data science publications rather than an exhaustive review of all publications. Overall, 8682 abstracts were screened, and 162 studies were included in this review (see Table, Supplemental Digital Content 2, which breaks down inclusion/exclusion numbers by outcome). Each of these studies were analyzed to identify their aims, study designs, data sources, samples, settings, populations, operational definitions of outcomes, list of variables, and data science methods.

RESULTS

Artificial Intelligence/Machine Learning Acceptance

Key Findings—Researchers approached the topics of Artificial Intelligence/Machine Learning (AI/ML) acceptance or credibility by measuring different outcomes. The selected papers investigated acceptance [4], satisfaction [5], trust [6], and use of AI[7]. Methodologically, two of the studies were quantitative [4,7], one qualitative [6], and one a mixed-method approach [5]. Finally, most of the selected research focused on specific AI-based products, such as a smartphone app [6], self-driving cars [4], and home assistants like Amazon's Echo [7].

Discussion—While the work of the Shin group is not specific to nursing[5], their conceptual model investigates algorithms based on the concepts of fairness, accountability, and transparency (FAT). These FAT principles are easily transferrable to the healthcare context. Subsequent studies should specifically investigate the relationship of the FAT concepts to credibility or acceptance in nursing-related AI/ML.

Burnout

Key Findings—All reports were cohort studies using survey-based data to predict some component of burnout. Three studies used logistic regression and one study used structural equation modeling (SEM) with path analysis. Notenbomer[8] explored absenteeism as a component of burnout, where both number of days and length of absence were reported. Bosman[9] similarly explored risk of sick leave, predicting this outcome as a binary variable within a timeframe. Oliver[10] explored subjective well-being in staff who care for those with intellectual disabilities, using the Satisfaction with Life Scale as a measure of well-being. And finally, Dutra[11] explored burnout using the Maslach Burnout inventory, with data collected from nurses and nursing technicians to formulate a predictive model.

Discussion—It is worth noting the limited number of studies in 2019 that discussed prediction around burnout and their variant approaches to measuring this phenomenon. Only one article approached burnout directly, using an established scale as the predictor. This

variance might be due to the fact that "burnout" as a term is not clearly delineated and has many aspects that could partially be defined. Here, we included both caregiver and healthcare professional burnout, but conceivably, burnout outside of the healthcare space could be examined, as it is a factor that influences individual health.

Of interest, the data contained in these studies were all collected using surveys and questionnaires. Often when considering data science methods, either real-time or historical data collected using a standardized method, e.g., the Electronic Health Record (EHR) or wearables, are used, and primary data collection is infrequent. It's possible the data required to predict burnout is not readily available, leading to a lack of more advanced data science methodology. In this light, we should promote the regular collection of data on staff and caregiver well-being, as we would then be able to develop decision-support tools to aid in minimizing the acquisition and effects of burnout. This is especially important as we approach increasing pressures regarding staffing shortages and costs associated with job attrition.

Emergency Department Visits

Key Findings—Screening of 626 studies resulted in 17 studies that met the inclusion criteria. Researchers have attempted to predict ED diagnoses such as sepsis[12], traumatic brain injury[13], and non-ST-elevation myocardial infarction patients[14]. Additional outcomes include risk prediction for opioid overdose[15], falls after ED discharge[16], urgent revascularization[17], ED readmission[18-21] and patient severity and eventual discharge disposition[22-25]. In addition to the traditional machine learning approaches, ED researchers also used NLP and social network analysis to predict patient outcomes. In two studies, NLP was used to analyze computed tomography, scan reports for prediction of subdural hematoma[26], and triage notes to predict discharge disposition[27]. Leone et al.[28] used an innovative application of social network analysis to classify women presenting to the ED due to violence exposure.

Discussion—A number of the studies highlighted the importance of nursing-collected information by exclusively or mostly using nursing triage data for prediction[12,18,22,24,25,27].

Falls

Key Findings—Researchers studied falls using both ML methods[16,29,30] and general predictive models employing various techniques such as least absolute shrinkage and selection operator (LASSO) to determine most relevant and important risk factors for falls[31-34]. The grave consequences of falls in terms of morbidity and mortality as well as healthcare costs have prompted many healthcare organizations to employ fall risk assessment scales, such as the Morse Fall Scale (MFS), used in two studies; Falls Risk Assessment Scoring System (FRASS), used in one study; and Functional Independence Measure (FIM), used in one study; as standard assessments of newly admitted patients to their settings or client caseloads. However, the uncertain value for appropriate fall risk classifications of patients has been an inducement for using the new, more sophisticated analysis methods to determine the assessment elements most important for risk mitigation. These goals appear to be foundational to the studies noted above that used predictive models

and spanned a variety of venues including tertiary care, rehabilitation, subacute, ambulatory care, and home health settings.

Discussion—The need to obtain thorough and accurate descriptions of fall episodes was highlighted in the publication sample as well, with Klock et al.[29] choosing to use ML to discern a real-time scoring method based on Agency for Healthcare Research and Quality (AHRQ) rubrics that can improve the quality of actual reports of fall incidents. These enhanced accounts can then contribute to more precise identification of valuable preventive interventions and elimination of ineffective practices.

Healthcare-Acquired Infections

Key Findings—Three studies used logistic regression to develop predictive models for the risk of healthcare acquired infection (HAI)-related outcomes. Hur et al.[35] developed a risk score for Catheter-Associated Urinary Tract Infections (CAUTI) that was incorporated into the EHR patient summary screen. Jackson et al.[36] developed a mixed model with better predictive performance than nares culture in identifying risk of nursing home residents transmitting methicillin-resistant *Staphylococcus aureus* (MRSA) to healthcare-worker gowns. They also used decision-curve analysis to compare the clinical utility of placing patients on contact precautions under each model. Lodise et al.[37] developed a bedside tool to predict the likelihood of six phenotypes of drug-resistant pathogens among hospitalized adult patients with Gram-negative infections. Their six logistic regression models were converted to an Excel-based user interface to estimate the risk of resistance at the bedside.

Four studies used more contemporary data science methods to predict HAI-related outcomes. Kocbek et al.[38] incorporated temporal data and preoperative blood tests to develop and compare four models to predict the onset of surgical site infection (SSI): several regression models and an extreme gradient boosting (XGBoost) model. Yee et al.[39] used publicly available ICU data and a data-driven approach using Bayesian networks and regression to develop a screening algorithm for progression into septic shock. Bush et al.[40] developed novel patient mobility predictors for unit-wide *Clostridium difficile* infection (CDI) susceptibility by using a network analysis and trace-route mapping to develop an in-hospital patient mobility network. The resulting calculated contagion centrality (CC) measure was found to be a statistically significant predictor of hospital-onset CDI cases. Liao et al.[41] used Cyranose 320 e-nose sensor breath-gas data to develop models of Ventilator-associated pneumonia (VAP) with *Pseudomonas aeruginosa* by applying neural network and support vector machine (SVM) methods. According to Liao and colleagues, while the combination of sensor data and ML methods shows promise, advancements in sensor performance and high-performance computing are needed to improve the accuracy of patient breathing gas detection models.

Discussion—System and unit-level implications of these findings suggest that addition of these types of predictors may enhance the development of more robust real-time surveillance systems for HAIs[40]. From a practice perspective, Hur et al.[35] specifically noted that a CAUTI surveillance system could reduce nurse time and effort used for risk assessment that could then be re-directed to management and education of indwelling urinary catheters.

From a research perspective, models that perform similarly and also offer lower complexity and higher interpretability are often preferred[38].

Healthcare Utilization and Costs

Key Findings—All studies explored different facets of costs and utilization, using several methods to do so. Five studies used NLP to explore outcomes[42-46], six used quantitative ML methods[47-52], two evaluated existing ML-based tools[53,54], and two used a form of regression analysis[55,56]. Interesting non-traditional data source examination was reported: video data with AI[57], deep learning with images[58], deep learning with audio data[59], and spatial analysis[60,61]. Outcomes reported were of such disparate interests as the prediction of financial risk of hospitalized pediatric patients[52], the identification of problematic opioid[46] use, and the evaluation of health literacy[43].

Discussion—The subject of "healthcare costs and utilization" covers a wide variety of topics and methods, and this is clearly reflected in the sample of papers included in this review. It is promising to see that several non-traditional data types (audio, image, text, geospatial, and video) are being used to the benefit of patient outcomes, reducing costs and increasing healthcare access beyond that of which traditional data are capable. A mix of direct and indirect economic-based outcomes were noted as well, including the use of deep learning-based image analysis to increase diagnostic quality of lower-dose positron emission tomography (PET) images[58], both reducing costs and advocating for patient safety. Noted throughout the studies, the purpose of cost analysis focused on some portion of patient advocacy, whether relating to cost-saving measures, treatment adherence, or clinical safety.

Hospitalization

Key Findings—A few themes emerged in these ten submissions. Data sources generally originated from existing administrative, commercial claim, and hospital data. Retrospective studies were a commonly adopted study design. Predictive and associative modeling dominate the data science methods employed in these studies. Several data modeling methods include risk prediction algorithm development[62], linear regression[63], multivariate statistical analysis using structural equation modeling[64], multivariable logistic regression[62,63,65-68], negative binomial-logit hurdle regression[69], geospatial analytic methods[60], and a network approach[40]. Ages and gender varied as did disease conditions. Financial impacts and implications appeared to be a common interest of study.

Discussion—The abundance of results and great variety of interests in leveraging data science methods to build predictive associations and relationships among different factors and variables pertaining to hospitalization are notable. The research in this space is showing promising results in mining predictive factors and associations to improve disease prevention and management, health promotion, and detecting gaps in geographical regions that relate to the impacts associated with hospitalization.

In-Hospital Mortality

Key Findings—A number of predictive models exist for identifying patients at high risk for dying in the hospital. The majority of the works used regression (with or

without additional methods) for making predictions[24,70-89]. The regression models primarily leveraged logistic regression; however, two papers applied Cox proportional hazards regression[83,87]. Ten papers noted the use of more contemporary methods for prediction: random forests[12,24,75,77,78,90], gradient boosting[12,24,75,78,83,91], Naïve Bayes[78], support vector machines[12,90], and neural networks[24,73,77,90]. Interestingly, one paper conducted a network analysis of healthcare providers and used the network characteristics to serve as predictors[81]. Another paper used regular expressions to extract features for a prediction model[77].

Sample sizes ranged from 51 to 281,522. Study populations included hospitalized adults from the following countries: Australia[70], Brazil[71], China[86,87,89], Israel[91], Ireland[70], Italy[84], Korea[76,77], Singapore[12], Spain[74], Switzerland[79], and the United States of America[24,72,75,78,80-82]. Several studies focused on specific admission diagnoses or surgical procedures, which resulted in a trend toward better model performance compared to models including all-cause hospitalizations. Variables serving as predictors primarily comprised: demographic information, vital signs, laboratory values, and diagnoses/comorbidities/ procedures. Less commonly included but notable predictor variables comprised: physical assessments[70], physiological status scores[71,74,75,81,91], and medication exposures[78,81,91]. One study included a nutrition score[71], one study included census-tract-level socioeconomic status[80], and one study included nursing diagnoses[84].

Discussion—All papers were limited to adult populations. There might be a need for pediatric-focused in-hospital mortality prediction models. From a nursing perspective, it was nice to see one paper include nursing diagnoses[84] and another paper include socioeconomic status[80]. These voids suggest promising areas for the nurse-investigator who possesses data science methods expertise or who works on the appropriately prepared interprofessional research team.

Length of Stay

Key Findings—Data science methods such as ML models (e.g., artificial neural network, predictive regression analysis) were used in five studies to project the hospital length of stay in different patient populations: (a) surgical patients undergoing orthopedic and neurosurgical operations[48,92,93], (b) patients who underwent surgeries as first-case in a day[94], and (c) critical care patients[95]. In two studies, NLP was used to characterize variables using narrative clinical notes[96] or patient comments[97] in order to study their association with hospital length of stay in a population of children with psychiatric complaints and orthopedic surgical patients, respectively.

Discussion—Only in one study[96] was nursing data (specifically, clinical notes written by triage or bedside nurses) used as predictors of length of stay even though several studies have shown the predictive power of nursing data on this outcome[98]. Further studies should include nursing data in predictive analytics methods to improve the prediction for patient and process outcomes.

Pain

Key Findings—Specific to predictive modeling, some studies focused on pain management[99-102] while other studies dealt with identifying predictors of pain[103-110]. There are studies that used biometric data to predict pain using brain grey-matter images[111] and another study that investigated the role of brain patterns in pain management among ED nurses[100]. Lee et al.'s[112] work using multi-modal imaging and autonomic signals as ML predictive variables is a very remarkable contribution in this area. Similarly, Jiang et al.[113] used multiple physiological parameters, galvanic skin response, and electromyography to predict pain level among non-verbal patients while Lim et al.[114] used a deep learning method on photoplethysmography signals to assess pain during surgery.

Discussion—The data sources used in the studies are mostly from the EHR, registry, public database, and clinical trial database containing data on inpatient/surgical encounters, survey questionnaires, and direct observation. Data types range from patient questionnaires (reported pain using numeric scale) to discrete observation data to more advanced biomedical measurements such as imaging data (e.g., magnetic resonance imaging) and physiologic sensor data (e.g., electromyography, photoplethysmography signals). These data were collected in various settings including hospitals, ambulatory clinics, community health, residential care, dental clinics, and sports medicine centers.

Data science techniques included least absolute shrinkage and selection operator, random forest regression, linear support vector classification, NLP of non-structured text, univariate/multivariate logistic regression, analysis of variance, cox proportional hazard regression model, k-means clustering, artificial neural network, support vector machine, multiple learning kernel regression model, multi-layer perceptron neural network, and deep belief network. Validation and testing techniques include bootstrapping, cluster sampling, leave-one-subject-out cross-validation, and 10-fold cross validation.

The use of data science in pain management has a major impact on medical and nursing practice. Its use can improve the ability to classify/measure pain in non-verbal patients (e.g., those on intra-operative deep sedation, those with altered levels of consciousness) and could change the way pain assessment is regarded as a purely subjective type of assessment. Advancing pain measurement or pain prediction based on various factors can greatly enhance the current pain management approach, which has the potential to reduce opioid overdose through personalized medicine. Prognostic pain models for chronic pain management can lead to more accurate opioid prescriptions for long-term opioid therapy, which may also reduce opioid overdose. Predictive models focusing on non-pharmacological interventions to pain management can improve medication safety by focusing on alternative options.

Patient Safety

Key Findings—The primary outcomes were the identification and classification of falls and fall incident reports[115-117], safety, and predicting perspectives of patient safety on the Hospital Survey on Patient Safety Culture[118]. Studies using data science techniques primarily used NLP[29,115,119], neural networks[29,120], and random forest[29,116] algorithms for

analysis. The data used to create models came from the EHR, narrative notes, and various surveys on hospital unit metrics. One study used online drug reference guides as part of a model to identify the potential for drug-drug interactions from EHR data[117].

Discussion—Only one patient safety study using data science techniques included nursing data[118]. No other studies directly used nursing data or were published in nursing journals. The limited exposure of nurses to data science techniques that investigate patient safety may be due to the lack of nursing researchers with expertise in patient safety and the use of data science techniques to create understanding.

Pressure Injuries

Key Findings—Shi, Dumville, and Cullum[121] reviewed, through a systematic review and meta-analysis, clinical effects of 22 prognostic models for predicting pressure injury (PI) risk; most of the models were built by logistic regression and Cox regression.

The seven empirical studies published in 2019 used various data science methods to detect or predict PIs. Li, Lin, and Hwang[122] explored several data mining algorithms to identify the best predictive factors on the occurrence of PIs, including logistic regression and three data mining algorithms (decision trees, neural networks, and support vector machines). The predictive factors, in order of importance, comprised: PI history, without cancer, excretion, activity/mobility, and skin condition/circulation.

Logistic regression analysis was used to compare three predictive models for PI occurrence in surgical patients[123] and to determine the utility of three different PI risk assessment scales (i.e., the Spinal Cord Injury Pressure Ulcer Scale, Braden Scale, and Functional Independence Measure [FIM]) for identifying individuals at risk for developing PI during inpatient spinal cord injury patients[124]. Park et al.[123] found that the Scott Triggers tool was the best fitting model; the estimated surgery time and serum albumin level were significant to predict the development of PIs in surgical patients in acute care settings. Flett et al.[124] found that the FIM bed/chair transfer score could be readily determined at rehabilitation admission with minimal administrative and clinical burden.

Crane et al.[125] used nonlinear regression to explore predictors for identifying patients at high risk of PI and recommended the need for modification of the Glamorgan scoring system, and incorporation of the Pediatric Logistic Organ Dysfunction-2 score might improve the predictive value of a modified Glamorgan scoring system. Zhang, Yu, Shi, Shang, Hong, and Yu[126] used a univariate Cox regression analysis to find the prognostic factors for PI recurrence. Blood albumin level on admission below 25 g/dl contributed to the strongest predicting factor for recurrence, followed by multiple ulcers and presence of a single caregiver. The variable of old age might constitute a risk factor for the pressure ulcer occurrence but not a prognostic factor. Duvall, Karg, Brienza, and Pearlman[127], using a threshold-based detection algorithm and a K-nearest neighbor classification approach, investigated the feasibility of a sensor technology (i.e., the E-scale system) for detecting and classifying movements in bed (i.e., roll, turn in place, extremity movements, and assisted turn), which are relevant for PI risk assessment. Ohura et al.[128] explored different architectures of the convolutional neural network (CNN) in image segmentation to detect

and discriminate ulcer regions of PI during assessment via telemedicine. The U-Net CNN constructed using appropriately supervised data was capable of segmentation with high accuracy. The study findings suggested that eHealth wound assessment using CNNs would be of practical use in the future.

Discussion—Data science methods facilitate the prediction, detection, and management of PIs via optimized assessments. Consideration of the best prognostic factors driven from the studies, such as blood albumin level, mobility, skin conditions, and single caregiver, can be used to develop and improve nutrition programs or home care nursing programs. Notably, nurses can improve their real-time monitoring of high-pressure areas in the bed and assessing PI risk with the use of sensor technologies (e.g., E-scale system). Also, as explored by Ohura et al.[128], the use of CNN architectures could support the eHealth wound assessment system to significantly change the management of PIs or chronic wounds. For future research, Park et al.[123] (2019) recommended the inclusion of vital signs and nursing interventions in PI predictive modeling.

Readmissions

Key Findings—The majority of papers used a readmission metric defined as unplanned readmission to an acute care hospital within 30 days of discharge[21,88,91,93,129-156], although some studies used 90-day[157,158], 180-day[159], within 1 year[160], 3 or more readmissions over 1 year[161], and even "instantaneous hospital readmission risk over time"[162]. Readmission was also defined by urgency[133,145] and etiology (e.g., disease-specific[129,143,144,150,159-161]). Given varying definitions, a paper by Brittan et al.[131] calculated three definitions of readmission with differing inclusion/exclusion criteria for index admissions and readmissions.

The populations studied were primarily adult; however, three papers focused on inpatients <18 years of age[21,131,148] or special populations such as surgical[88,93,134,137,139-142,153,156-158] and trauma[160], medical conditions such as heart failure[129,143,144,155,161], heart failure or myocardial infarction[159], COPD[135], diabetes[142], IV[150], falls[160], antimicrobial therapy[134], and skilled nursing facility discharge[132,163]. The majority of papers applied logistic regression in prediction models with 22 papers using development and validation cohorts, boot-strapped internal validation, or cross validation[21,91,93,129,131-134,136,138,140-144,147,150,151,155,159,161,162].

One paper applied principal components analysis, multiple correspondence analysis, and multiple factor analysis[129]. Eleven papers applied machine learning algorithms such as random forest, weighted decision trees, support vector machines, gradient boosting, neural networks, decision curve analysis, SMOTE[91,93,129,130,132,142-144,147,148], and Naïve-Bayes[21]. In one nurse-authored paper by Kwon et al.[142] a case study was used to illustrate different statistical and ML risk models and hospital readmission outcomes of patients with diabetes mellitus.

Novel predictors used in risk models with relevance to nursing included tests of physical function[132,135,137,139,145,153-156,158,160], symptoms[132,135,139,158,159,161], psychosocial factors[150,154,155,157,160,161], vital

signs and/or body mass index[88,91,129,135,137,138,143,144,148,152,156-158,160,161], and frailty[88,135,137,139,145,153,155,156,158,160]. Other predictors that are infrequently applied in prevailing risk models include laboratory and/or imaging tests[88,91,129,132,138,140,142,144,148,150-153,155,157,159,161,164], and medications[91,136,138,142,146,148,150,151,154,157,160,161]. Nihhawan et al.[150] incorporated novel socio-behavioral predictors such as health literacy, adherence to medications, substance abuse, patient-provider relationship satisfaction, perceived health status, and housing and food security.

Discussion—Most models used administrative claims and EHR data for sociodemographic information and medical diagnoses. Overall, multiple papers demonstrated that risk factors such as older age, poor health, frailty, multimorbidity, certain medical diagnoses, and healthcare utilization confer high risk for readmission. While these risk factors potentially improve the predictive ability of models, nurses can make important contributions to model development by filling data gaps with nursing-relevant data pertaining to patients' biopsychosocial health and function. Identifying and applying common data elements relevant to nursing across EHR systems in predictive models and including standard nursing terminology (e.g., "International Classification of Nursing Practice" codes in the EHR as suggested by Kwon et al.[142]) would capture some nuances that provide contextual information about patient health status and thereby improve the relevance and performance of the models.

Staffing/Scheduling/Workload

Key Findings—Data science methods were reportedly used in three studies involving an estimation of the antecedents or consequences of nurse staffing that year. Two studies reported predictive models formed from ML methods, and one study was a report of NLP used to transform clinical notes into assessment forecasting. The Nadkarni[165] research group used a stepwise, iterative, object-oriented program written with workflow and treatment processes in mind in a sample of 343 patients with potentially life-threatening complications and 2,285 uncomplicated mothers in a Tanzanian hospital. Aimed at providing decision-makers with a tool to analyze the impact of resource limitations on maternal inpatient complications, key variables included treatment efficacy, severity distribution, number and frequency of nurse visits, nurse staffing at the shift level, deterioration rate, and maternal near-misses.

Similarly, the Lucero[34] group (2019) elucidated a data-driven and practice-based approach to identify factors associated with inpatient falls in a sample of 272 patients who fell and 542 who did not while hospitalized in medical-surgical units of a Florida tertiary-care hospital. Manual, semi-automated, and automated procedures deploying theoretically or practice-derived risk factors yielded a meaningful and parsimonious set of predictors for this adverse event. Skill mix, rates of nurse certification, and nurse-educational levels were among the relevant staffing variables in this observational case-controlled study.

In the remaining Menger[166] study, NLP was used to transform clinical notes from the patient's EHR to develop and validate a multivariable prediction model for the assessment

of inpatient violence risk. In this prognostic study, the authors used clinician notes from the admission encounters of over 5,000 patients in one of two different psychiatric settings in The Netherlands. The model training and estimation of predictive validity was done in a nested cross-validation setup in which the outcome of interest—the manifestation of violent behavior within four weeks of admission—was successfully predicted from inpatient violence risk assessment derived from the documentation in this manner. Although a staffing variable was not explicitly or operationally stated, the availability of a nurse (or psychiatrist) to conduct the admission assessment is inferred in this initial encounter from which language within the nursing (and medical) domain is derived.

Discussion—No articles related to *Scheduling* or *Workload* were found. In these studies we reviewed, staffing variables were of two types: nurse hours relative to either all staffing or patient load as well as nurse characteristics such as education/certification. Further, studies of the impact nurse staffing may have on patient outcomes should include characteristics of the nurse which are known or hypothesized to have an impact, such as their education, training, and mentoring needs. From a systems perspective, measures of the human capital resources, e.g. nurse hours/patient day or skill mix, should be explicitly stated and for the relevant time partitions up to and including the time of injury, adversity, or other measurement.

Unit Culture

Of the 589 papers yielded in the initial literature search, none of the studies satisfied criteria for being included in the final analysis.

DISCUSSION

Through our literature review, we have identified and described a representative sample of publications focused on the use of data science methods relevant to nurses. All but one of the outcomes for which we searched were associated with at least one paper published in 2019. From a methodological perspective, we noted the use of many contemporary data science methods (e.g., natural language processing, neural networks, and social network analysis) throughout many of the outcomes.

We found a large number of studies exploring *Readmissions* and *Pressure Injuries* (PI). Risk prediction modeling for hospital readmission has increased in recent years due to the Affordable Care Act of 2010 and the subsequent Hospital Readmission Reduction Program which has tied financial reimbursement penalties to potentially avoidable hospital readmissions. The high number of PI studies could be attributable to either (a) PI risk scores have existed for many years, so there is ample opportunity for including validated predictors within new analysis frameworks, and/or (b) PIs are regulatory quality indicators associated with malpractice litigation and excess costs.

Conversely, several topics (i.e., *Artificial Intelligence/Machine Learning Acceptance, Burnout, Patient Safety,* and *Unit Culture*) were poorly represented and could be areas where there is an opportunity to leverage data science methods in research on these nursing topics. In fact, our *Unit Culture* search did not reveal any results. While this could be a limitation

of our search strategy, it is worth considering that more studies could be performed in this space. Given nurses' long-standing attention to these latter areas of *Burnout, Patient Safety,* and *Unit Culture*, we are hopeful the nursing research and nursing informatics communities will apply data science methods to these problems in the coming years. Additionally, we believe identifying and applying common data elements relevant to nursing across EHR systems in predictive models and including standard nursing terminology codes in the EHR would capture some nuances that provide contextual information about patient health status. The inclusion of these codes could be worth pursuing in future research efforts, as it could be high-yield.

Limitations of our report include the non-exhaustive nature of the literature search and the single-person review process. Given that the intent of the paper was to provide readers with a broad overview of nursing-relevant data science activities, an exhaustive literature search was beyond our purpose. For interested readers, we have published search strategies so that others can reproduce our findings and/or perform an exhaustive literature review. The use of a single-person review helped expedite the process of a year-in-review paper. Additionally, because we are focused on high-level description rather than inferential comparisons, the use of a second reviewer would not have significantly changed our findings.

CONCLUSION

Data science has significant potential to assist healthcare providers in improving the nursing environment, clinical processes, and patient outcomes. By using data science techniques to identify care environment improvement opportunities and/or individual patient risk factors, we create new opportunities to design and implement interventions best able to mitigate risk and improve patient care. The use of data science to understand problems related to nursing and nursing care must include modern methods of investigation and understanding. We hope the studies and reports we have identified and described in this paper will help readers understand the breadth and depth of data science's ability to improve clinical processes and patient outcomes that are relevant to nurses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Center for Nursing Informatics. Data Science Workgroup Paper presented at: Nursing Knowledge: Big Data Science Conference 2019; Minneapolis, MN.

2. Heslop L, Lu S, Xu X. Nursing-sensitive indicators: a concept analysis. J Adv Nurs. 2014;70(11):2469–2482. [PubMed: 25113388]

3. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Systematic reviews. 2016;5(1):210. [PubMed: 27919275]

4. Liu P, Yang R, Xu Z. Public Acceptance of Fully Automated Driving: Effects of Social Trust and Risk/Benefit Perceptions. Risk Anal. 2019;39(2):326–341. [PubMed: 30059602]

5. Shin D, Park YJ. Role of fairness, accountability, and transparency in algorithmic affordance. Computers in Human Behavior. 2019;98:277–284.

6. Lee Y, Ha M, Kwon S, Shim Y, Kim J. Egoistic and altruistic motivation: How to induce users' willingness to help for imperfect AI. Computers in Human Behavior. 2019;101:180–196.

7. McLean G, Osei-Frimpong K. Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. Computers in Human Behavior. 2019;99:28–37.

8. Notenbomer A, van Rhenen W, Groothoff JW, Roelen CAM. Predicting long-term sickness absence among employees with frequent sickness absence. International archives of occupational and environmental health. 2019;92(4):501–511. [PubMed: 30474733]

9. Bosman LC, Roelen CAM, Twisk JWR, Eekhout I, Heymans MW. Development of Prediction Models for Sick Leave Due to Musculoskeletal Disorders. Journal of occupational rehabilitation. 2019;29(3):617–624. [PubMed: 30607694]

10. Oliver A, Revuelto L, Fernandez I, Simo-Algado S, Galiana L. An integrative model of the subjective well-being of staff working in intellectual disability services. Res Dev Disabil. 2019;87:1–8. [PubMed: 30685682]

11. Dutra HS, Guirardello EB, Li Y, Cimiotti JP. Nurse Burnout Revisited: A Comparison of Computational Methods. J Nurs Meas. 2019;27(1):E17–E33. [PubMed: 31068498]

12. Chiew CJ, Liu N, Tagami T, Wong TH, Koh ZX, Ong MEH. Heart rate variability based machine learning models for risk prediction of suspected sepsis patients in the emergency department. Medicine (Baltimore). 2019;98(6):e14197. [PubMed: 30732136]

13. Bertsimas D, Dunn J, Steele DW, Trikalinos TA, Wang Y. Comparison of Machine Learning Optimal Classification Trees With the Pediatric Emergency Care Applied Research Network Head Trauma Decision Rules. JAMA Pediatr. 2019;173(7):648–656. [PubMed: 31081856]

14. Wu CC, Hsu WD, Islam MM, et al. An artificial intelligence approach to early predict non-ST-elevation myocardial infarction patients with chest pain. Comput Methods Programs Biomed. 2019;173:109–117. [PubMed: 31046985]

15. Lo-Ciganic WH, Huang JL, Zhang HH, et al. Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions. JAMA Netw Open. 2019;2(3):e190968. [PubMed: 30901048]

16. Patterson BW, Engstrom CJ, Sah V, et al. Training and Interpreting Machine Learning Algorithms to Evaluate Fall Risk After Emergency Department Visits. Med Care. 2019;57(7):560–566. [PubMed: 31157707]

17. Goto S, Kimura M, Katsumata Y, et al. Artificial intelligence to predict needs for urgent revascularization from 12-leads electrocardiography in emergency patients. PLoS One. 2019;14(1):e0210103. [PubMed: 30625197]

18. Bergese I, Frigerio S, Clari M, et al. An Innovative Model to Predict Pediatric Emergency Department Return Visits. Pediatr Emerg Care. 2019;35(3):231–236. [PubMed: 27741066]

19. Howell P, Elkin PL. Can Solo Practitioners Survive in Value-Based Healthcare? Validating a Predicative Model for ED Utilization. Studies in health technology and informatics. 2019;264:1682–1683. [PubMed: 31438291]

20. Vest JR, Ben-Assuli O. Prediction of emergency department revisits using area-level social determinants of health measures and health information exchange information. International journal of medical informatics. 2019;129:205–210. [PubMed: 31445257]

21. Wolff P, Grana M, Rios SA, Yarza MB. Machine Learning Readmission Risk Modeling: A Pediatric Case Study. BioMed research international. 2019;2019:8532892. [PubMed: 31139655]

22. Goto T, Camargo CA Jr., Faridi MK, Freishtat RJ, Hasegawa K. Machine Learning-Based Prediction of Clinical Outcomes for Children During Emergency Department Triage. JAMA Netw Open. 2019;2(1):e186937. [PubMed: 30646206]

23. Kramer J, Schreyogg J, Busse R. Classification of hospital admissions into emergency and elective care: a machine learning approach. Health Care Manag Sci. 2019;22(1):85–105. [PubMed: 29177993]

24. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA Jr., Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Critical care (London, England). 2019;23(1):64.

25. Rendell K, Koprinska I, Kyme A, Ebker-White AA, Dinh MM. The Sydney Triage to Admission Risk Tool (START2) using machine learning techniques to support disposition decision-making. Emerg Med Australas. 2019;31(3):429–435. [PubMed: 30469164]

26. Pruitt P, Naidech A, Van Ornam J, Borczuk P, Thompson W. A natural language processing algorithm to extract characteristics of subdural hematoma from head CT reports. Emerg Radiol. 2019;26(3):301–306. [PubMed: 30693414]

27. Sterling NW, Patzer RE, Di M, Schrager JD. Prediction of emergency department patient disposition based on natural language processing of triage notes. International journal of medical informatics. 2019;129:184–188. [PubMed: 31445253]

28. Leone M, Lapucci E, De Sario M, Davoli M, Farchi S, Michelozzi P. Social network analysis to characterize women victims of violence. BMC Public Health. 2019;19(1):494. [PubMed: 31046717]

29. Klock M, Hong K, Yang G. Scoring Patient Fall Reports Using Quality Rubric and Machine Learning…The 17th World Congress of Medical and Health Informatics, 25-30 August 2019, Lyon, France. Studies in Health Technology & Informatics. 2019;264:639–643. [PubMed: 31438002]

30. Lo Y, Lynch SF, Urbanowicz RJ, et al. Using Machine Learning on Home Health Care Assessments to Predict Fall Risk…The 17th World Congress of Medical and Health Informatics, 25-30 August 2019, Lyon, France. Studies in Health Technology & Informatics. 2019;264:684–688. [PubMed: 31438011]

31. Fusco-Gessick B, Cournan M. Using Functional Independence Measure Subscales to Predict Falls-Rapid Assessment. Rehabilitation nursing : the official journal of the Association of Rehabilitation Nurses. 2019;44(4):236–244. [PubMed: 29557822]

32. Oshiro CES, Frankland TB, Rosales AG, et al. Fall Ascertainment and Development of a Risk Prediction Model Using Electronic Medical Records. Journal of the American Geriatrics Society. 2019;67(7):1417–1422. [PubMed: 30875089]

33. Than S, Crabtree A, Moran C. Examination of risk scores to better predict hospital-related harms. Intern Med J. 2019;49(9):1125–1131. [PubMed: 30270479]

34. Lucero RJ, Lindberg DS, Fehlberg EA, et al. A data-driven and practice-based approach to identify risk factors associated with hospital-acquired falls: Applying manual and semi- and fully-automated methods. International journal of medical informatics. 2019;122:63–69. [PubMed: 30623785]

35. Hur EY, Jin Y, Jin T, Lee SM. Development and Evaluation of the Automated Risk Assessment System for Catheter-Associated Urinary Tract Infection. Computers, informatics, nursing : CIN. 2019;37(9):463–472.

36. Jackson SS, Lydecker AD, Magder LS, Roghmann MC. Development and Validation of a Clinical Prediction Rule to Predict Transmission of Methicillin-Resistant Staphylococcus aureus in Nursing Homes. Am J Epidemiol. 2019;188(1):214–221. [PubMed: 30351349]

37. Lodise TP, Bonine NG, Ye JM, Folse HJ, Gillard P. Development of a bedside tool to predict the probability of drug-resistant pathogens among hospitalized adult patients with gram-negative infections. BMC Infect Dis. 2019;19(1):718. [PubMed: 31412809]

38. Kocbek P, Fijacko N, Soguero-Ruiz C, et al. Maximizing Interpretability and Cost-Effectiveness of Surgical Site Infection (SSI) Predictive Models Using Feature-Specific Regularized Logistic Regression on Preoperative Temporal Data. Comput Math Methods Med. 2019;2019:2059851. [PubMed: 30915154]

39. Yee CR, Narain NR, Akmaev VR, Vemulapalli V. A Data-Driven Approach to Predicting Septic Shock in the Intensive Care Unit. Biomed Inform Insights. 2019;11:1178222619885147.

40. Bush K, Barbosa H, Farooq S, et al. Predicting hospital-onset Clostridium difficile using patient mobility data: A network approach. Infection Control & Hospital Epidemiology. 2019;40(12):1380–1386. [PubMed: 31656216]

41. Liao YH, Wang ZC, Zhang FG, Abbod MF, Shih CH, Shieh JS. Machine Learning Methods Applied to Predict Ventilator-Associated Pneumonia with Pseudomonas aeruginosa Infection via Sensor Array of Electronic Nose in Intensive Care Unit. Sensors (Basel). 2019;19(8).

42. Skaljic M, Patel IH, Pellegrini AM, Castro VM, Perlis RH, Gordon DD. Prevalence of Financial Considerations Documented in Primary Care Encounters as Identified by Natural Language Processing Methods. JAMA Network Open. 2019;2(8):e1910399–e1910399. [PubMed: 31469397]

43. Balyan R, Crossley SA, Brown W 3rd, et al. Using natural language processing and machine learning to classify health literacy from secure messages: The ECLIPPSE study. PLoS One. 2019;14(2):e0212488. [PubMed: 30794616]

44. Tang Y, Yang J, Ang PS, et al. Detecting adverse drug reactions in discharge summaries of electronic medical records using Readpeer. International journal of medical informatics. 2019;128:62–70. [PubMed: 31160013]

45. Bigeard E, Thiessard F, Grabar N. Detecting Drug Non-Compliance in Internet Fora Using Information Retrieval and Machine Learning Approaches. Studies in health technology and informatics. 2019;264:30–34. [PubMed: 31437879]

46. Masters ET, Ramaprasan A, Mardekian J, et al. Natural Language Processing-Identified Problem Opioid Use and Its Associated Health Care Costs. J Pain Palliat Care Pharmacother. 2018;32(2-3):106–115. [PubMed: 30702378]

47. Liu L, Ni Y, Zhang N, Nick Pratap J. Mining patient-specific and contextual data with machine learning technologies to predict cancellation of children's surgery. International journal of medical informatics. 2019;129:234–241. [PubMed: 31445261]

48. Ramkumar PN, Navarro SM, Haeberle HS, et al. Development and Validation of a Machine Learning Algorithm After Primary Total Hip Arthroplasty: Applications to Length of Stay and Payment Models. The Journal of arthroplasty. 2019;34(4):632–637. [PubMed: 30665831]

49. Hussain OA, Junejo KN. Predicting treatment outcome of drug-susceptible tuberculosis patients using machine-learning models. Inform Health Soc Care. 2019;44(2):135–151. [PubMed: 29461901]

50. Sun C, Ippel L, van Soest J, et al. A Privacy-Preserving Infrastructure for Analyzing Personal Health Data in a Vertically Partitioned Scenario. Studies in health technology and informatics. 2019;264:373–377. [PubMed: 31437948]

51. Taranik M, Kopanitsa G. Using Machine Learning for Personalized Patient Adherence Level Determination. Studies in health technology and informatics. 2019;261:174–178. [PubMed: 31156111]

52. Lin E, Hefner JL, Zeng X, et al. A deep learning model for pediatric patient risk stratification. The American journal of managed care. 2019;25(10):e310–e315. [PubMed: 31622071]

53. Bundorf MK, Polyakova M, Stults C, et al. Machine-Based Expert Recommendations And Insurance Choices Among Medicare Part D Enrollees. Health Affairs. 2019;38(3):482–490. [PubMed: 30830808]

54. Forman EM, Goldstein SP, Zhang F, et al. OnTrack: development and feasibility of a smartphone app designed to predict and prevent dietary lapses. Translational behavioral medicine. 2019;9(2):236–245. [PubMed: 29617911]

55. Kan HJ, Kharrazi H, Chang HY, Bodycombe D, Lemke K, Weiner JP. Exploring the use of machine learning for risk adjustment: A comparison of standard and penalized linear regression models in predicting health care costs in older adults. PLoS One. 2019;14(3):e0213258. [PubMed: 30840682]

56. Desai R, Jo A, Marlow NM. Risk for Medication Nonadherence Among Medicaid Enrollees With Fibromyalgia: Development of a Validated Risk Prediction Tool. Pain Pract. 2019;19(3):295–302. [PubMed: 30369018]

57. Xiong GL, Bayen E, Nickels S, et al. Real-time video detection of falls in dementia care facility and reduced emergency care. The American journal of managed care. 2019;25(7):314–315. [PubMed: 31318502]

58. Kaplan S, Zhu Y-M. Full-Dose PET Image Estimation from Low-Dose PET Image Using Deep Learning: a Pilot Study. Journal of Digital Imaging. 2019;32(5):773–778. [PubMed: 30402670]

59. Monge-Alvarez J, Hoyos-Barcelo C, Lesso P, Casaseca-de-la-Higuera P. Robust Detection of Audio-Cough Events Using Local Hu Moments. IEEE J Biomed Health Inform. 2019;23(1):184–196. [PubMed: 29994432]

60. Delamater PL, Shortridge AM, Kilcoyne RC. Using floating catchment area (FCA) metrics to predict health care utilization patterns. BMC Health Serv Res. 2019;19(1):144. [PubMed: 30832628]

61. Ramirez AG, Schneider EB, Mehaffey JH, Zeiger MA, Hanks JB, Smith PW. Effect of Travel Time for Thyroid Surgery on Treatment Cost and Morbidity. Am Surg. 2019;85(9):949–955. [PubMed: 31638505]

62. Chang ET, Piegari R, Wong ES, et al. Which Patients Are Persistently High-Risk for Hospitalization? American Journal of Managed Care. 2019;25(9):e274–e281.

63. Arora S, Stouffer GA, Kucharska-Newton AM, et al. Twenty Year Trends and Sex Differences in Young Adults Hospitalized With Acute Myocardial Infarction. Circulation. 2019;139(8):1047–1056. [PubMed: 30586725]

64. Kattan WM, Abduljawad AA. Predicting different factors that affect hospital utilization and outcomes Among diabetic patients admitted with hypoglycemia using structural equation modeling. Diabetes research and clinical practice. 2019;153:55–65. [PubMed: 31152808]

65. Lim S, Gangoli G, Adams E, et al. Increased Clinical and Economic Burden Associated With Peripheral Intravenous Catheter-Related Complications: Analysis of a US Hospital Discharge Database. Inquiry. 2019;56:46958019875562. [PubMed: 31524024]

66. Bhavsar GP, Probst JC, Bennett KJ, Hardin JW, Qureshi Z. Community-level electronic prescribing and adverse drug event hospitalizations among older adults. Health Informatics Journal. 2019;25(3):661–675. [PubMed: 28737062]

67. Hsu DY, Smith B, Silverberg JI. Atopic Dermatitis and Hospitalization for Mental Health Disorders in the United States. Dermatitis. 2019;30(1):54–61. [PubMed: 30570573]

68. Vanderlaan J, Rochat R, Williams B, Dunlop A, Shapiro SE. Associations Between Hospital Maternal Service Level and Delivery Outcomes. Women's health issues : official publication of the Jacobs Institute of Women's Health. 2019;29(3):252–258.

69. Stockbridge EL, Chhetri S, Polcar LE, Loethen AD, Carney CP. Behavioral health conditions and potentially preventable diabetes-related hospitalizations in the United States: Findings from a national sample of commercial claims data. PLoS One. 2019;14(2):e0212955. [PubMed: 30818377]

70. Cardona M, Lewis E, Shanmugam S, et al. Dissonance on perceptions of end-of-life needs between health-care providers and members of the public: Quantitative cross-sectional surveys. Australas J Ageing. 2019;38(3):e75–e84. [PubMed: 30868725]

71. de Lima Junior JD, Matias JEF, Stahlke Junior HJ. Risk factors associated with hospital mortality in mitral valve reoperation. Rev Col Bras Cir. 2019;46(3):e20192176. [PubMed: 31432982]

72. Haider A, Con J, Prabhakaran K, et al. Developing a Simple Clinical Score for Predicting Mortality and Need for ICU in Trauma Patients. Am Surg. 2019;85(7):733–737. [PubMed: 31405418]

73. Hassanipour S, Ghaem H, Arab-Zozani M, et al. Comparison of artificial neural network and logistic regression models for prediction of outcomes in trauma patients: A systematic review and meta-analysis. Injury. 2019;50(2):244–250. [PubMed: 30660332]

74. Herran-Monge R, Muriel-Bombin A, Garcia-Garcia MM, et al. Epidemiology and Changes in Mortality of Sepsis After the Implementation of Surviving Sepsis Campaign Guidelines. J Intensive Care Med. 2019;34(9):740–750. [PubMed: 28651474]

75. Hill BL, Brown R, Gabel E, et al. An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data. Br J Anaesth. 2019;123(6):877–886. [PubMed: 31627890]

76. Kim Y, Kym D, Hur J, et al. Development of a risk prediction model (Hangang) and comparison with clinical severity scores in burn patients. PLoS One. 2019;14(2):e0211075. [PubMed: 30726241]

77. Kwon JM, Kim KH, Jeon KH, Park J. Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography. Echocardiography. 2019;36(2):213–218. [PubMed: 30515886]

78. Lei VJ, Kennedy EH, ThaiBinh L, et al. Model Performance Metrics in Assessing the Value of Adding Intraoperative Data for Death Prediction: Applications to Noncardiac Surgery…The 17th World Congress of Medical and Health Informatics, 25-30 August 2019, Lyon, France. Studies in Health Technology & Informatics. 2019;264:223–227. [PubMed: 31437918]

79. Malinovska A, Pitasch L, Geigy N, Nickel CH, Bingisser R. Modification of the Emergency Severity Index Improves Mortality Prediction in Older Patients. The western journal of emergency medicine. 2019;20(4):633–640. [PubMed: 31316703]

80. McCarthy FH, Zhang L, Tam V, et al. Geographically Derived Socioeconomic Factors to Improve Risk Prediction in Patients Having Aortic Valve Replacement. Am J Cardiol. 2019;123(1):116–122. [PubMed: 30390990]

81. Park Y, Karampourniotis PD, Sylla I, Das AK. Hierarchical patient-centric caregiver network method for clinical outcomes study. PLoS One. 2019;14(2):e0211218. [PubMed: 30759091]

82. Pickens RC, Bloomer AK, Sulzer JK, et al. Modifying Interhospital Hepatopancreatobiliary Transfers Based on Predictive Analytics: Moving from a Center of Excellence to a Health-Care System of Excellence. Am Surg. 2019;85(9):1033–1039. [PubMed: 31638520]

83. Pieszko K, Hiczkiewicz J, Budzianowski P, et al. Predicting Long-Term Mortality after Acute Coronary Syndrome Using Machine Learning Techniques and Hematological Markers. Dis Markers. 2019;2019:9056402. [PubMed: 30838085]

84. Sanson G, Welton J, Vellone E, et al. Enhancing the performance of predictive models for Hospital mortality by adding nursing data. International journal of medical informatics. 2019;125:79–85. [PubMed: 30914184]

85. Waller A, Turon H, Bryant J, Zucca A, Evans TJ, Sanson-Fisher R. Medical oncology outpatients' preferences and experiences with advanced care planning: A cross-sectional study. BMC Cancer. 2019;19(1).

86. Wang L, Yang F, Wang X, et al. Predicting mortality in patients undergoing VA-ECMO after coronary artery bypass grafting: the REMEMBER score. Critical care (London, England). 2019;23(1):11.

87. Wu L, Zhang Z, Wang Y, et al. A Model to Predict In-Hospital Mortality in HIV/AIDS Patients with Pneumocystis Pneumonia in China: The Clinical Practice in Real World. BioMed research international. 2019;2019:6057028. [PubMed: 30906778]

88. Zhang S, Tan S, Jiang Y, et al. Sarcopenia as a predictor of poor surgical and oncologic outcomes after abdominal surgery for digestive tract cancer: A prospective cohort study. Clinical Nutrition. 2019;38(6):2881–2888. [PubMed: 30630709]

89. Zhang J, Cheng B, Yang M, Pan J, Feng J, Cheng Z. Predicting in-hospital death in patients with type B acute aortic dissection. Medicine (Baltimore). 2019;98(32):e16462. [PubMed: 31393350]

90. Lin K, Hu Y, Kong G. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. International journal of medical informatics. 2019;125:55–61. [PubMed: 30914181]

91. Zeltzer D, Balicer RD, Shir T, Flaks-Manov N, Einav L, Shadmi E. Prediction Accuracy With Electronic Medical Records Versus Administrative Claims. Med Care. 2019;57(7):551–559. [PubMed: 31135691]

92. Danilov G, Kotik K, Shifrin M, Strunina U, Pronkina T, Potapov A. Prediction of Postoperative Hospital Stay with Deep Learning Based on 101 654 Operative Reports in Neurosurgery. Studies in health technology and informatics. 2019;258:125–129. [PubMed: 30942728]

93. Merrill RK, Ferrandino RM, Hoffman R, Shaffer GW, Ndu A. Machine Learning Accurately Predicts Short-Term Outcomes Following Open Reduction and Internal Fixation of Ankle Fractures. J Foot Ankle Surg. 2019;58(3):410–416. [PubMed: 30803914]

94. Tafti AP, Dong Y, Habermann E, Liu H, Herasevich V. Relationship Between Very Cold Outside Weather and Surgical Outcome: Integrating Shallow and Deep Artificial Neural Nets. Studies in health technology and informatics. 2019;264:1783–1784. [PubMed: 31438342]

95. Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. Sci Data. 2019;6(1):96. [PubMed: 31209213]

96. McCoy TH, Wiste AK, Doyle AE, Pellegrini AM, Perlis RH. Association between child psychiatric emergency room outcomes and dimensions of psychopathology. General Hospital Psychiatry. 2019;59:1–6. [PubMed: 31034963]

97. Menendez ME, Shaker J, Lawler SM, Ring D, Jawa A. Negative Patient-Experience Comments After Total Shoulder Arthroplasty. The Journal of bone and joint surgery American volume. 2019;101(4):330–337. [PubMed: 30801372]

98. Sanson G, Vellone E, Kangasniemi M, Alvaro R, D'Agostino F. Impact of nursing diagnoses on patient and organisational outcomes: a systematic literature review. J Clin Nurs. 2017;26(23-24):3764–3783. [PubMed: 28042921]

99. Parthipan A, Banerjee I, Humphreys K, et al. Predicting inadequate postoperative pain management in depressed patients: A machine learning approach. PLoS One. 2019;14(2):e0210575. [PubMed: 30726237]

100. Corradi-Dell'Acqua C, Foerster M, Sharvit G, et al. Pain management decisions in emergency hospitals are predicted by brain activity during empathy and error monitoring. Br J Anaesth. 2019;123(2):e284–e292. [PubMed: 30916038]

101. Oosterhaven J, Wittink H, Dekker J, Kruitwagen C, Deville W. Pain catastrophizing predicts dropout of patients from an interdisciplinary chronic pain management programme: A prospective cohort study. J Rehabil Med. 2019;51(10):761–769. [PubMed: 31544215]

102. Gilpin HR, Stahl DR, McCracken LM. A theoretically guided approach to identifying predictors of treatment outcome in Contextual Cognitive Behavioural Therapy for chronic pain. European journal of pain (London, England). 2019;23(2):354–366.

103. Wang M, Li Y, Li J, Fan L, Yu H. The risk of moderate-to-severe post-operative pain following the placement of dental implants. J Oral Rehabil. 2019;46(9):836–844. [PubMed: 31074878]

104. Kanbayashi Y, Sakaguchi K, Nakatsukasa K, et al. Predictive factors for taxane acute pain syndrome determined by ordered logistic regression analysis. Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer. 2019;27(7):2673–2677. [PubMed: 30478672]

105. Pua YH, Poon CL, Seah FJ, et al. Predicting individual knee range of motion, knee pain, and walking limitation outcomes following total knee arthroplasty. Acta Orthop. 2019;90(2):179–186. [PubMed: 30973090]

106. Larsson B, Dragioti E, Gerdle B, Bjork J. Positive psychological well-being predicts lower severe pain in the general population: a 2-year follow-up study of the SwePain cohort. Ann Gen Psychiatry. 2019;18(1):8. [PubMed: 31164910]

107. Cashin AG, Traeger AC, Hubscher M, et al. Persistent Pain After Wrist or Hand Fracture: Development and Validation of a Prognostic Model. The Journal of orthopaedic and sports physical therapy. 2019;49(1):28–35. [PubMed: 30208793]

108. Priego Quesada JI, Kerr ZY, Bertucci WM, Carpes FP. A retrospective international study on factors associated with injury, discomfort and pain perception among cyclists. PLoS One. 2019;14(1):e0211197. [PubMed: 30682136]

109. Hah JM, Cramer E, Hilmoe H, et al. Factors Associated With Acute Pain Estimation, Postoperative Pain Resolution, Opioid Cessation, and Recovery: Secondary Analysis of a Randomized Clinical Trial. JAMA Netw Open. 2019;2(3):e190168. [PubMed: 30821824]

110. Page MG, Boyd K, Ware MA. Examination of the Course of Low Back Pain Intensity Based on Baseline Predictors and Health Care Utilization Among Patients Treated in Multidisciplinary Pain Clinics: A Quebec Pain Registry Study. Pain Med. 2019;20(3):564–573. [PubMed: 30690515]

111. Chen T, Mu J, Xue Q, et al. Whole-brain structural magnetic resonance imaging-based classification of primary dysmenorrhea in pain-free phase: a machine learning study. Pain. 2019;160(3):734–741. [PubMed: 30376532]

112. Lee J, Mawla I, Kim J, et al. Machine learning-based prediction of clinical pain using multimodal neuroimaging and autonomic metrics. Pain. 2019;160(3):550–560. [PubMed: 30540621]

113. Jiang M, Mieronkoski R, Syrjala E, et al. Acute pain intensity monitoring with the classification of multiple physiological parameters. J Clin Monit Comput. 2019;33(3):493–507. [PubMed: 29946994]

114. Lim H, Kim B, Noh GJ, Yoo SK. A Deep Neural Network-Based Pain Classifier Using a Photoplethysmography Signal. Sensors (Basel). 2019;19(2).

115. Chapman AB, Peterson KS, Alba PR, DuVall SL, Patterson OV. Detecting Adverse Drug Events with Rapidly Trained Classification Models. Drug Saf. 2019;42(1):147–156. [PubMed: 30649737]

116. Gupta J, Patrick J, Poon S. Clinical Safety Incident Taxonomy Performance on C4.5 Decision Tree and Random Forest. Studies in health technology and informatics. 2019;266:83–88. [PubMed: 31397306]

117. Shiima Y, Wong ZS. Classification Scheme for Incident Reports of Medication Errors. Studies in health technology and informatics. 2019;265:113–118. [PubMed: 31431586]

118. Stimpfel AW, Djukic M, Brewer CS, Kovner CT. Common predictors of nurse-reported quality of care and patient safety. Health care management review. 2019;44(1):57–66. [PubMed: 28263207]

119. Liu J, Wong ZS, Tsui KL, So HY, Kwok A. Exploring Hidden In-Hospital Fall Clusters from Incident Reports Using Text Analytics. Studies in health technology and informatics. 2019;264:1526–1527. [PubMed: 31438214]

120. Bouzillé G, Morival C, Westerlynck R, et al. An Automated Detection System of Drug-Drug Interactions from Electronic Patient Records Using Big Data Analytics…The 17th World Congress of Medical and Health Informatics, 25-30 August 2019, Lyon, France. Studies in Health Technology & Informatics. 2019;264:45–49. [PubMed: 31437882]

121. Shi C, Dumville JC, Cullum N. Evaluating the development and validation of empirically-derived prognostic models for pressure ulcer risk assessment: A systematic review. Int J Nurs Stud. 2019;89:88–103. [PubMed: 30352322]

122. Li HL, Lin SW, Hwang YT. Using Nursing Information and Data Mining to Explore the Factors That Predict Pressure Injuries for Patients at the End of Life. Computers, informatics, nursing : CIN. 2019;37(3):133–141.

123. Park SK, Park HA, Hwang H. Development and Comparison of Predictive Models for Pressure Injuries in Surgical Patients: A Retrospective Case-Control Study. J Wound Ostomy Continence Nurs. 2019;46(4):291–297. [PubMed: 31274856]

124. Flett HM, Delparte JJ, Scovil CY, Higgins J, Laramee MT, Burns AS. Determining Pressure Injury Risk on Admission to Inpatient Spinal Cord Injury Rehabilitation: A Comparison of the FIM, Spinal Cord Injury Pressure Ulcer Scale, and Braden Scale. Arch Phys Med Rehabil. 2019;100(10):1881–1887. [PubMed: 31054293]

125. Crane N, Pool N, Chang I, Rogan S, Stocker C, Raman S. A dedicated paediatric logistic organ dysfunction score - adjusted pressure injury risk assessment scale is required for tertiary paediatric ICUs. Cardiol Young. 2019;29(3):455–456. [PubMed: 30678745]

126. Zhang N, Yu X, Shi K, Shang F, Hong L, Yu J. A retrospective analysis of recurrent pressure ulcer in a burn center in Northeast China. J Tissue Viability. 2019;28(4):231–236. [PubMed: 31405543]

127. Duvall J, Karg P, Brienza D, Pearlman J. Detection and classification methodology for movements in the bed that supports continuous pressure injury risk assessment and repositioning compliance. J Tissue Viability. 2019;28(1):7–13. [PubMed: 30598376]

128. Ohura N, Mitsuno R, Sakisaka M, et al. Convolutional neural networks for wound detection: the role of artificial intelligence in wound care. J Wound Care. 2019;28(Sup10):S13–S24.

129. Mahajan SM, Ghani R. Combining Structured and Unstructured Data for Predicting Risk of Readmission for Heart Failure Patients. Studies in health technology and informatics. 2019;264:238–242. [PubMed: 31437921]

130. Baig M, Hua N, Zhang E, et al. Predicting Patients at Risk of 30-Day Unplanned Hospital Readmission. Studies in health technology and informatics. 2019;266:20–24. [PubMed: 31397296]

131. Brittan MS, Campagna EJ, Keller D, Kempe A. How Measurement Variability Affects Reporting of a Single Readmission Metric. Journal for Healthcare Quality: Promoting Excellence in Healthcare. 2019;41(3):160–164.

132. Chandra A, Rahman PA, Sneve A, et al. Risk of 30-Day Hospital Readmission Among Patients Discharged to Skilled Nursing Facilities: Development and Validation of a Risk-Prediction Model. Journal of the American Medical Directors Association. 2019;20(4):444–444. [PubMed: 30852170]

133. Deschepper M, Eeckloo K, Vogelaers D, Waegeman W. A hospital wide predictive model for unplanned readmission using hierarchical ICD data. Comput Methods Programs Biomed. 2019;173:177–183. [PubMed: 30777619]

134. Durojaiye OC, Kritsotakis EI, Johnston P, Kenny T, Ntziora F, Cartwright K. Developing a risk prediction model for 30-day unplanned hospitalization in patients receiving outpatient parenteral antimicrobial therapy. Clin Microbiol Infect. 2019;25(7):905 e901–905 e907. [PubMed: 30502491]

135. Ehsani H, Mohler MJ, Golden T, Toosizadeh N. Upper-extremity function prospectively predicts adverse discharge and all-cause COPD readmissions: a pilot study. Int J Chron Obstruct Pulmon Dis. 2019;14:39–49. [PubMed: 30587960]

136. Franckowiak TM, Raub JN, Yost R. Derivation and validation of a hospital all-cause 30-day readmission index. American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists. 2019;76(7):436–443. [PubMed: 31361822]

137. Holzgrefe RE, Wilson JM, Staley CA, Anderson TL, Wagner ER, Gottschalk MB. Modified frailty index is an effective risk-stratification tool for patients undergoing total shoulder arthroplasty. Journal of shoulder and elbow surgery. 2019;28(7):1232–1240. [PubMed: 30878278]

138. Kabue S, Greene J, Kipnis P, et al. The Impact of Pharmacy-specific Predictors on the Performance of 30-Day Readmission Risk Prediction Models. Med Care. 2019;57(4):295–299. [PubMed: 30829940]

139. Katlic MR, Coleman J, Khan K, Wozniak SE, Abraham JH. Sinai Abbreviated Geriatric Evaluation: Development and Validation of a Practical Test. Ann Surg. 2019;269(1):177–183. [PubMed: 29189383]

140. Khera S, Kolte D, Deo S, et al. Derivation and external validation of a simple risk tool to predict 30-day hospital readmissions after transcatheter aortic valve replacement. EuroIntervention. 2019;15(2):155–163. [PubMed: 30803938]

141. Kim LD, Pfoh ER, Hu B, et al. Derivation and Validation of a Model to Predict 30-Day Readmission in Surgical Patients Discharged to Skilled Nursing Facility. J Am Med Dir Assoc. 2019;20(9):1086–1090 e1082. [PubMed: 31176675]

142. Kwon JY, Karim ME, Topaz M, Currie LM. Nurses "Seeing Forest for the Trees" in the Age of Machine Learning: Using Nursing Knowledge to Improve Relevance and Performance. Computers, informatics, nursing : CIN. 2019;37(4):203–212.

143. Mahajan SM, Ghani R. Using Ensemble Machine Learning Methods for Predicting Risk of Readmission for Heart Failure. Studies in health technology and informatics. 2019;264:243–247. [PubMed: 31437922]

144. Mahajan SM, Mahajan A, Burman P, Heidenreich P. Can We Do More With Less While Building Predictive Models? A Study in Parsimony of Risk Models for Predicting Heart Failure Readmissions. CIN: Computers, Informatics, Nursing. 2019;37(6):306–314.

145. McAlister F, van Walraven C. External validation of the Hospital Frailty Risk Score and comparison with the Hospital-patient One-year Mortality Risk Score to predict outcomes in elderly hospitalised patients: a retrospective cohort study. BMJ quality & safety. 2019;28(4):284–288.

146. McConachie SM, Raub JN, Trupianio D, Yost R. Development of an iterative validation process for a 30-day hospital readmission prediction index. American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists. 2019;76(7):444–452. [PubMed: 31361819]

147. Morgan DJ, Bame B, Zimand P, et al. Assessment of Machine Learning vs Standard Prediction Rules for Predicting Hospital Readmissions. JAMA Netw Open. 2019;2(3):e190348. [PubMed: 30848808]

148. Nakamura MM, Toomey SL, Zaslavsky AM, et al. Potential Impact of Initial Clinical Data on Adjustment of Pediatric Readmission Rates. Acad Pediatr. 2019;19(5):589–598. [PubMed: 30470563]

149. Brooke MS, Nakamura MU, Hosomi JK, Ribeiro MC, Sass N. Translation and validation of Warmometer, a tool for assessing warmth in patient-provider relationships, for use in Brazilian Portuguese. Sao Paulo Med J. 2018;136(3):192–199. [PubMed: 30020344]

150. Nijhawan AE, Metsch LR, Zhang S, et al. Clinical and Sociobehavioral Prediction Model of 30-Day Hospital Readmissions Among People With HIV and Substance Use Disorder: Beyond Electronic Health Record Data. J Acquir Immune Defic Syndr. 2019;80(3):330–341. [PubMed: 30763292]

151. Pauly V, Mendizabal H, Gentile S, Auquier P, Boyer L. Predictive risk score for unplanned 30-day rehospitalizations in the French universal health care system based on a medico-administrative database. PLoS One. 2019;14(3):e0210714. [PubMed: 30861004]

152. Robinson R, Bhattarai M, Hudali T. Vital Sign Abnormalities on Discharge Do Not Predict 30-Day Readmission. Clin Med Res. 2019;17(3-4):63–71. [PubMed: 31324735]

153. Shahrokni A, Tin A, Alexander K, et al. Development and Evaluation of a New Frailty Index for Older Surgical Patients With Cancer. JAMA Netw Open. 2019;2(5):e193545. [PubMed: 31074814]

154. Sieck C, Adams W, Burkhart L. Validation of the BOOST Risk Stratification Tool as a Predictor of Unplanned 30-Day Readmission in Elderly Patients. Quality Management in Health Care. 2019;28(2):96–102. [PubMed: 30921282]

155. Sokoreli I, Cleland JG, Pauws SC, et al. Added value of frailty and social support in predicting risk of 30-day unplanned re-admission or death for patients with heart failure: An analysis from OPERA-HF. Int J Cardiol. 2019;278:167–172. [PubMed: 30587417]

156. Traven SA, Reeves RA, Sekar MG, Slone HS, Walton ZJ. New 5-Factor Modified Frailty Index Predicts Morbidity and Mortality in Primary Hip and Knee Arthroplasty. The Journal of arthroplasty. 2019;34(1):140–144. [PubMed: 30337252]

157. Goltz DE, Ryan SP, Hopkins TJ, et al. A Novel Risk Calculator Predicts 90-Day Readmission Following Total Joint Arthroplasty. The Journal of bone and joint surgery American volume. 2019;101(6):547–556. [PubMed: 30893236]

158. Woldu SL, Sanli O, Clinton TN, Lotan Y. Validating the predictors of outcomes after radical cystectomy for bladder cancer. Cancer. 2019;125(2):223–231. [PubMed: 30291813]

159. Cediel G, Sandoval Y, Sexter A, et al. Risk Estimation in Type 2 Myocardial Infarction and Myocardial Injury: The TARRACO Risk Score. The American journal of medicine. 2019;132(2):217–226. [PubMed: 30419227]

160. Hatcher VH, Galet C, Lilienthal M, Skeete DA, Romanowski KS. Association of Clinical Frailty Scores With Hospital Readmission for Falls After Index Admission for Trauma-Related Injury. JAMA Netw Open. 2019;2(10):e1912409. [PubMed: 31577357]

161. Wang N, Gallagher R, Sze D, Hales S, Tofler G. Predictors of Frequent Readmissions in Patients With Heart Failure. Heart, lung & circulation. 2019;28(2):277–283.

162. Chen S, Kong N, Sun X, Meng H, Li M. Claims data-driven modeling of hospital time-to-readmission risk with latent heterogeneity. Health Care Manag Sci. 2019;22(1):156–179. [PubMed: 29372450]

163. Attar HS, Chandramani S. Impact of physician empathy on migraine disability and migraineur compliance. Ann Indian Acad Neurol. 2012;15(Suppl 1):S89–94. [PubMed: 23024571]

164. Mahajan NN, Gaikwad NL, Mahajan KN, Soni RN. Internal iliac artery ligation for arresting postpartum haemorrhage [1]. BJOG: An International Journal of Obstetrics and Gynaecology. 2007;114(7):906. [PubMed: 17567423]

165. Nadkarni D, Minocha A, Harpaldas H, et al. Predicting resource-dependent maternal health outcomes at a referral hospital in Zanzibar using patient trajectories and mathematical modeling. PLoS One. 2019;14(3):e0212753. [PubMed: 30835755]

166. Menger V, Spruit M, van Est R, Nap E, Scheepers F. Machine Learning Approach to Inpatient Violence Risk Assessment Using Routinely Collected Clinical Notes in Electronic Health Records. JAMA Network Open. 2019;2(7):e196709–e196709. [PubMed: 31268542]
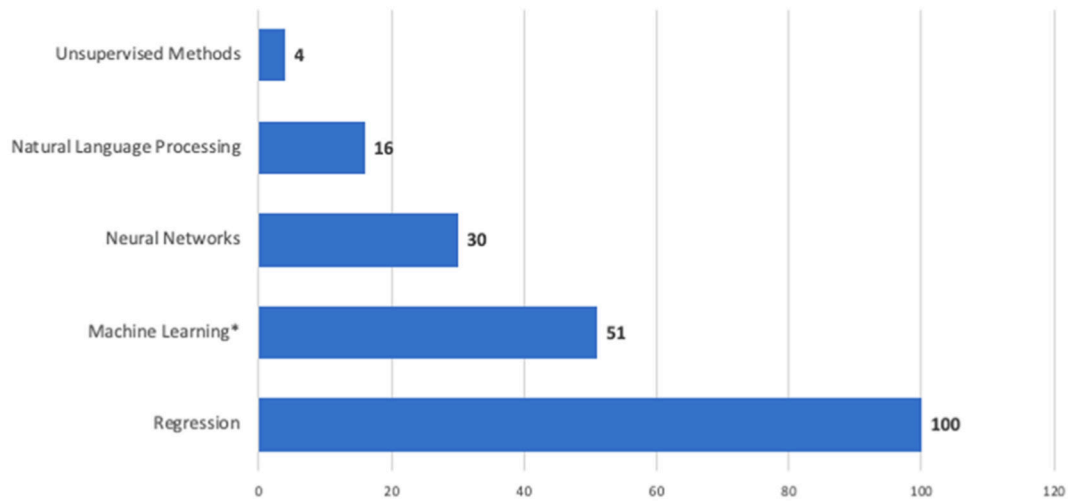
FIGURE 1.

Frequency of use of data science methods among reviewed studies. *Note: While several methods on this figure could be considered "machine learning," this count does not include studies counted in a different category.
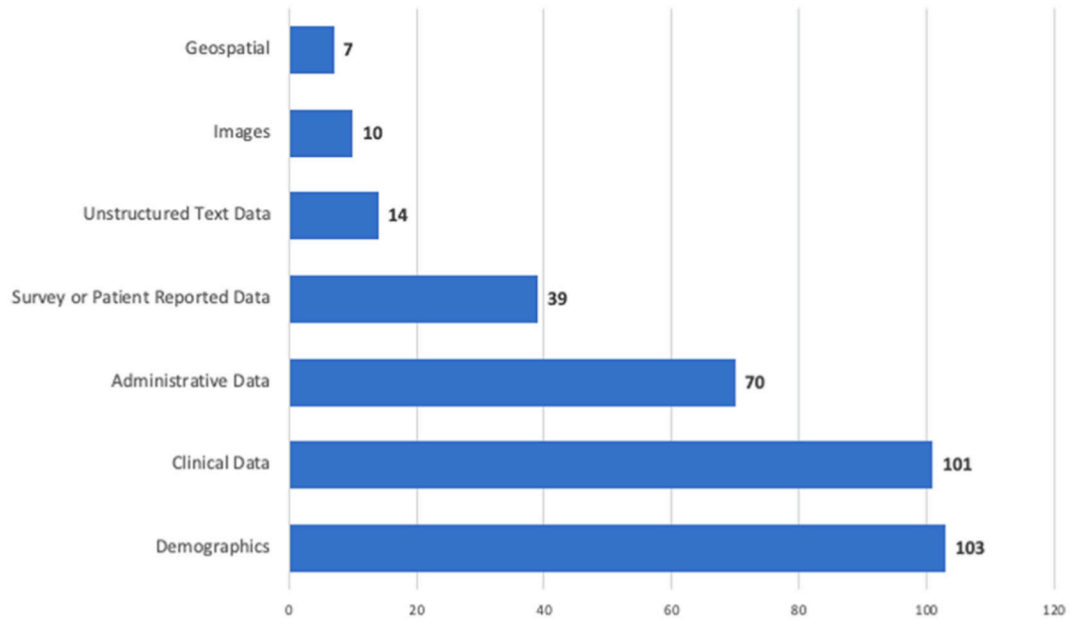
FIGURE 2.
Frequency of use of independent variable categories among reviewed studies.