ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Systematic Literature Review

A Scoping Review of the Use of Machine Learning in Health Economics and Outcomes Research: Part 2—Data From Nonwearables

Woojung Lee, PharmD, Naomi Schwartz, PhD, Aasthaa Bansal, PhD, Sara Khor, MASc, Noah Hammarlund, PhD, Anirban Basu, PhD, Beth Devine, PhD, PharmD, MBA

ABSTRACT

Objectives: Despite the increasing interest in applying machine learning (ML) methods in health economics and outcomes research (HEOR), stakeholders face uncertainties in when and how ML can be used. We reviewed the recent applications of ML in HEOR.

Methods: We searched PubMed for studies published between January 2020 and March 2021 and randomly chose 20% of the identified studies for the sake of manageability. Studies that were in HEOR and applied an ML technique were included. Studies related to wearable devices were excluded. We abstracted information on the ML applications, data types, and ML methods and analyzed it using descriptive statistics.

Results: We retrieved 805 articles, of which 161 (20%) were randomly chosen. Ninety-two of the random sample met the eligibility criteria. We found that ML was primarily used for predicting future events (86%) rather than current events (14%). The most common response variables were clinical events or disease incidence (42%) and treatment outcomes (22%). ML was less used to predict economic outcomes such as health resource utilization (16%) or costs (3%). Although electronic medical records (35%) were frequently used for model development, claims data were used less frequently (9%). Tree-based methods (eg, random forests and boosting) were the most commonly used ML methods (31%).

Conclusions: The use of ML techniques in HEOR is growing rapidly, but there remain opportunities to apply them to predict economic outcomes, especially using claims databases, which could inform the development of cost-effectiveness models.

Keywords: health data, health economics, machine learning, nonwearable data, outcomes research.

VALUE HEALTH. 2022; 25(12):2053–2061

Introduction

In recent years, interest in machine learning (ML) methods has increased in multiple research areas.[1,2] ML generally refers to algorithms that learn from data with minimal human assistance or intervention.[3,4] Such techniques have received attention because of their better capability of analyzing big and complicated data sets than traditional approaches.[4] Such abilities of ML techniques may enable researchers to tackle more novel and complex research questions where assumptions for traditional modeling approaches cannot be justified and give opportunities to use diverse types of data. Furthermore, thanks to the advancement of computational power, building highly advanced ML techniques (eg, super learners) has become easier than before, helping researchers resolve challenges that heretofore were beyond the capabilities of traditional modeling approaches.

Health economics and outcomes research (HEOR) is one of the research areas where ML can play an important role, considering a rapid increase in the availability of big health data.[5,6] Enormous volumes of health data are being derived from numerous sources, including more traditional sources such as electronic medical records (EMRs) and payer records and newer sources such as wearables, patient portals, and social media. With the advancement of generating and storing such data, health data are likely to become available in a higher volume and speed over time and richer in its content. Furthermore, several initiatives (eg, Optum Laps and the Patient-Centered Outcomes and Research Initiative Clinical Data Research Networks) have created large repositories for various types of health data, further advancing the generation and utilization of health data.[5,7]

Advanced ML techniques, accompanied by the growing volume and depth of health data, can expand the breadth of research topics in HEOR and provide new ways of tackling research questions.[8] In addition to using it to predict clinical health outcomes, researchers can make innovative use of ML techniques to predict economic outcomes such as healthcare utilization and costs of interventions, potentially affecting the coverage and pricing decisions.[9] Identifying heterogeneity in treatment response and health policy interventions is another research area in HEOR where ML can play a role, informing more personalized or targeted interventions.[10,11] These are just a few examples of the research areas in HEOR where researchers can benefit from using ML, and again, with the rapidly advancing techniques and data, the range of ML applications in HEOR will become more expansive in the coming years.

Despite the potential value of ML in HEOR and an increasing interest in using it among stakeholders, the adoption of ML has been nascent, and the role of ML in HEOR has been far less studied. A few studies provided a high-level overview of the potential applications of ML in overall health-related research, but none of them were HEOR-specific.[12,13] Therefore, we aimed to address the gap by reviewing how ML has been applied in HEOR in recent years and to discuss the potential opportunities of using ML in the field of HEOR. We separated the review into 2 parts, wearable specific and nonwearable specific, because of a particular interest in wearable data in HEOR and potential differences in the application of ML between the wearable data and the other types of health data. We presented the findings specific to wearable data in part 1. In part 2, we focus on nonwearable data.

Methods

Data Sources and Study Selection

We conducted a scoping review using PubMed (the International Prospective Register of Systematic Reviews registration number: CRD42021260881). Study inclusion criteria were studies that (1) were published between January 2020 and March 2021, (2) focused on HEOR, (3) applied an ML technique, and (4) used nonwearable data (we reviewed studies using wearable data in part 1 of this study). Studies were excluded if they were (1) nonoriginal studies (eg, review or commentary), (2) methods articles, (3) nonhuman studies, (4) non-English studies, (5) studies without full text available, and (6) studies using an image, video, or omics data only (this was to narrow the scope by excluding the types of data that were generally not used in HEOR).

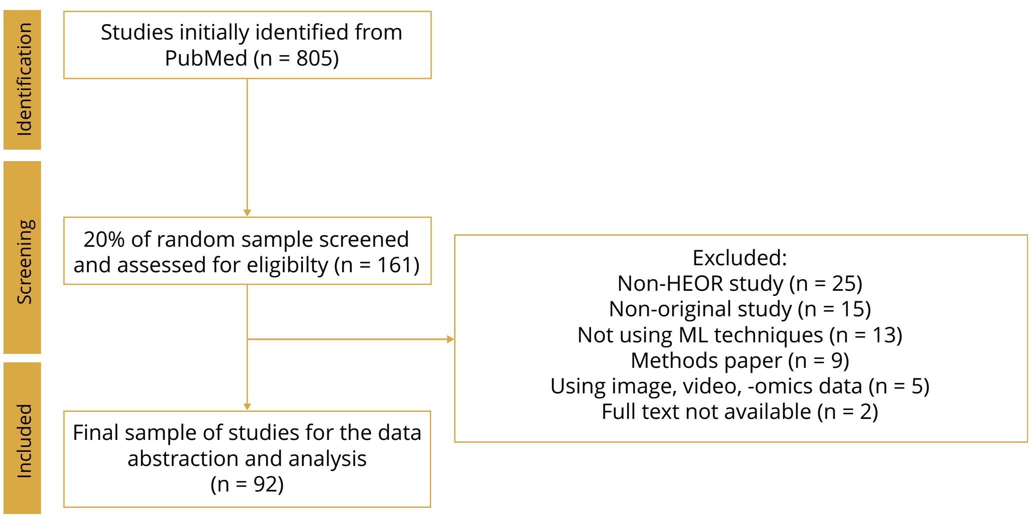
Given that the aim of this study was to examine the recent trends of ML use in HEOR, we included studies published from 2020. We defined HEOR as research focusing on the clinical and economic aspects of health or health interventions.[14] It was challenging to develop a list of search terms that are comprehensive enough to identify all HEOR studies because HEOR is a very broad and nebulous term. Therefore, we alternatively decided to use Medical Subject Headings (MeSH) related to HEOR. We reviewed all MeSH terms included under the healthcare category and chose those related to HEOR. Studies were required to have at least one HEOR MeSH term and satisfy the definition of HEOR to be considered as an HEOR study.

We developed search strategies with the support of an experienced systematic review librarian. The prespecified search terms are presented in Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.07.011. Once we identified studies from PubMed, we took a random sample of 20% of the studies because we judged that a sample of approximately 100 articles will provide sufficient coverage. Then, 2 reviewers (W.L. and N.S.) independently screened the title and abstract of the random-sampled studies based on the inclusion and exclusion criteria. The reviewers further reviewed the full text for uncertain studies. Conflicts in terms of the included articles and reasons for exclusion were resolved by the discussion of the 2 reviewers and the rest of the coauthors.

Data Abstraction

One reviewer (W.L.), in collaboration with the other authors, developed a standardized data abstraction that consisted of 4 parts. In part 1, we abstracted general information of each included study using the PICOTS (ie, patient population, intervention, comparator, outcome, timing, and setting) framework. In part 2, we collected data on (1) types of disease areas, (2) application purposes, (3) types of model outcomes, and (4) application settings. The categories for 1 to 3 were the same as the wearable-specific review study (ie, the part 1 of our study). We categorized application settings into trial versus real-world settings and generated subcategories for the real-world settings based on whether the study design was retrospective, prospective, or both. In part 3, we reviewed 4 items related to the data set used for ML models: (1) types of training data set, (2) types of testing data set, (3) the size of training data set, and (4) the size of testing data set. We prespecified types of data sets based on the existing studies that stratified the types of data used in HEOR: EMRs,

Figure 1. PRISMA diagram.



PRISMA indicates Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Table 1. Types of model outcomes by goal of application (n = 92).

Model outcomes	Goal of application			Examples for model outcomes	References
	Forecasting	Monitoring	All		
Clinical event or disease incidence	35 (38%)	4 (4%)	39 (42%)	The incidence of postpartum depression, the development of cardiac events, and death	17,25,27,31,34,35,37,39,42-45,52,56,59,60,62,63,65,69,70,73,75,76,79-81,83,84,88,90-96,100,104
Treatment outcomes	18 (20%)	2 (2%)	20 (22%)	RECIST criteria in patients treated with PD-1 blockade, complications after total hip arthroplasty, and outcomes for cognitive behavioral therapy	22,23,26,28,32,36,38,47,49,50,53-55,58,66,68,77,82,86,103
Healthcare resource utilization	15 (16%)	0 (0%)	15 (16%)	The use of prolonged mechanical ventilation, initiation of renal replacement therapy, and presurgery healthcare resource utilization	16,19,21,24,48,61,67,74,78,85,87,98,99,101,102
Disease progression and symptoms	3 (3%)	4 (4%)	7 (7%)	Functional level in subarachnoid hemorrhage, the level of persistent pain, and the progression of schizophrenia	29,46,57,64,71,89,97
Others	3 (2%)	2 (2%)	5 (4%)	Patient punctuality, awareness of precautionary procedures, and medication nonadherence risk	18,20,30,106,107
Costs	3 (3%)	0 (0%)	3 (3%)	Out-of-pocket health expenditures, healthcare cost under the Oncology Care Model, and costs with digital health platform	40,41,72
General health status	2 (2%)	1 (1%)	3 (3%)	General health status, multimorbidity frailty, and dependence in the activities of daily living	33,51,105
Total	79 (86%)	13 (14%)	92 (100%)		

Note: Values are presented as n (%).
PD-1 indicates programmed cell death protein 1; RECIST, response evaluation criteria in solid tumors.

administrative claims data, registry data, primary data, and others.[12,15] We defined primary data as data collected directly by researchers themselves. “Others” include any other data that were not part of the other 4 categories (eg, social media data) and linked data (eg, the Surveillance, Epidemiology, and End Results-Medicare). The training and testing data size was defined as the total number of observations in each of the data sets. Part 4 included the following items related to ML methods used: (1) types of ML techniques used, (2) the form of the outcome variables, and (3) the performance metrics reported. The definition and categories for each item were consistent with the wearable-specific review study. With regard to the types of ML techniques, we descriptively examined whether these techniques vary by the types of model outcomes (abstracted in part 1) and the types of the training data set (abstracted in part 3).

If a study satisfied multiple categories for a given item (eg, a study that used multiple types of ML methods or multiple performance metrics), we abstracted all of them and counted them separately. All items were described through frequency distributions.

Results

Reviewed Studies

Our search identified a total of 805 studies from PubMed, and we randomly chose to screen 161 of them (20%). Ninety-two studies met the eligibility criteria and thus were included in the study.[16-107] The most frequent reasons for exclusion included not being an HEOR study (n = 25), nonoriginal study (n = 15), not using ML techniques (n = 13), and methods article (n = 9) (Fig. 1). Included studies are presented in Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.07.011.

Applications of ML

Most of the included studies (74 of 92, 80%) were disease specific, with joint diseases (10 of 92, 11%) and cardiovascular diseases (10 of 92, 11%) being the most common types of disease where ML was applied (see Figure S1 in Appendix 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.07.011). Most studies (77 of 92, 83%) were conducted in the real-world setting, with most studies being retrospective (61 of 77, 86%). Notably, 5% of the included studies (5 of 92) were conducted in the clinical trial setting.

ML was more frequently used to forecast future outcomes (79 of 92, 86%) than to monitor current outcomes (13 of 92, 14%) (Table 1[16-107]). Clinical outcomes such as clinical events or disease incidence (39 of 92, 42%) and treatment outcomes (20 of 92, 22%) accounted for most types of model outcomes. Economic outcomes such as healthcare resource utilization (15 of 92, 16%) and costs (3 of 92, 3%) were relatively less frequently represented.

Types of Data Used With ML

The secondary use of EMR (33 of 92, 35%) and the primary data collection (22 of 92, 23%) were relatively frequently used. Claims data (8 of 92, 9%) were the type of data least frequently represented among all major data types. In Table 2,[16-107] we have listed the examples of specific data sets along with the frequency of use for each type of training data set. We observed a similar distribution for the types of testing data sets (data not shown).

EMR was the most used type of data when predicting clinical events or disease incidence, treatment outcomes, healthcare

Table 2. Frequency distribution of training data sets and examples of model outcomes (n = 92).

Types of training data sets	Examples of data sets	Examples of model outcomes	n (%)	References
EMR	Retrospective chart review, outpatient clinical data, and radiological data	The risk of postpartum depression, postoperative death, and admission to the intensive care unit	33 (35)	17,22-24,30,35,36,38,39,44,45,47,51,52,56,58,61,62,65,67,70,74,79,83,86,91,97,101-103,105
Primary data collection	A survey conducted in the outpatient clinic, daily outpatient blood collection, a survey conducted among wounded US service members	The demand for medical care, outcomes for cognitive behavioral therapy, and age-related macular degeneration	22 (23)	18,20,21,28,29,31,42,54,57,59,64,66,71,76,77,89,93-96,100,107
Registry data	NSQIP data set, OSHPD data set, and US department of defense trauma registry	The risk of prolonged medical ventilation, trauma patient mortality, and outcomes of limb revascularization	14 (15)	16,19,26,32,34,37,46,49,53,55,69,75,82,84
Claims data	IBM MarketScan, administrative data from an AMC in Singapore, and Medicare claims data	Presurgery healthcare resource utilization, persistent high utilizer, and postpartum psychiatric admission	8 (9)	33,48,72,78,85,87,88,90
Miscellaneous	Free-text pharmacy prescription databases, administrative data from clinics, ambulance calls, and twitter	Medication incidents, patient wait time, and demand for ambulances	16 (17)	25,27,40,41,43,50,60,63,68,73,81,92,98,99,104,106
Not used*	N/A	N/A	1 (1)	80

AMC indicates Academic Medical Center; EMR, electronic medical record; NSQIP, National Surgical Quality Improvement Program; N/A, not applicable; OSHPD, California Office of Statewide Health and Planning and Development.
*Validation study.

utilization, and general health status (Fig. 2). Primary data, in contrast, were dominantly used when measuring disease progression or symptoms, and claims data were mostly used to predict costs and healthcare resource utilization.

There were 9 studies (9%, 9 of 92) without a clear information about the size of the training data set. Among the remaining 83 studies, the median number of observations in the training data set was 2022 with lower (25%) and upper (75%) quartiles of 510 and 19 954, respectively. In terms of the testing data set, 23% of the studies (23 of 92) did not have an independent data set and used cross-validation to test the models. Twenty percent (20 of 92) did not specifically mention whether or how the model was tested.

Figure 2. Types of training data based on types of model outcomes. Note: others include biomarkers, patient punctuality, awareness of precautionary procedures, and therapeutic alliance; only 4 nonmiscellaneous types of data are shown In this figure, and thus, the numbers are smaller than what is presented in Tables 1 and 2.

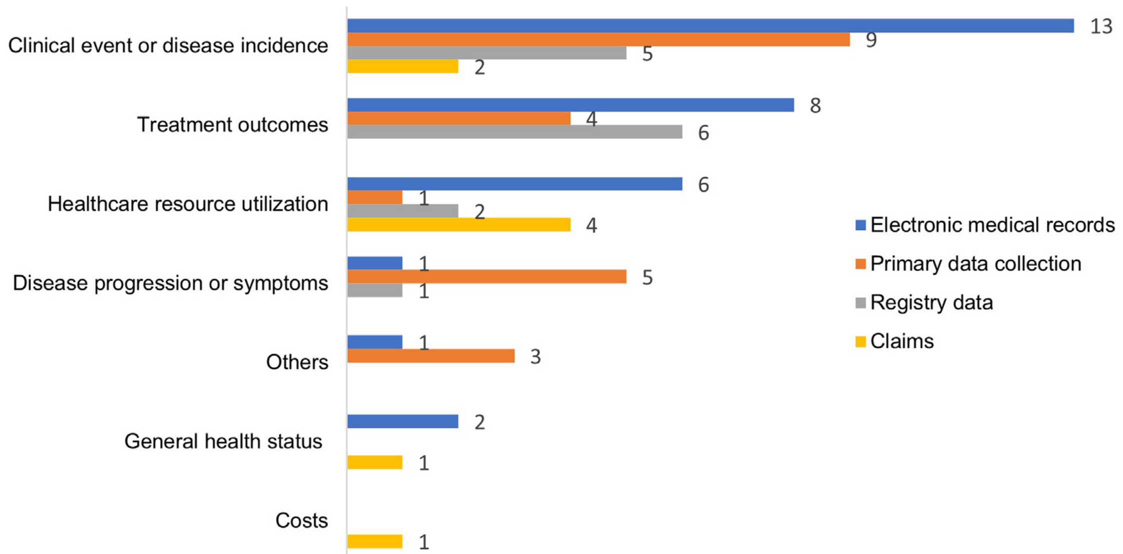


Table 3. Types of ML models based on types of applications (N = 210).*

Types of ML models	Goal of application		All	References
	Forecasting	Monitoring		
Tree-based models	59 (28%)	6 (3%)	65 (31%)	16,17,20,26-31,33,34,36-45,46,47,49-59,62,63,65,67-70,72, 74-85,87,91,92,94-96,98-100,103,105
Logistic/linear regression	36 (17%)	2 (1%)	38 (18%)	16,20,22,23,26,27,29-31,37-41,45,50,56,58,59,61,65, 67-69,72,74-77,82-84,87,90,91,101,102,105
Support vector machine	24 (11%)	6 (3%)	30 (14%)	16,20,23,25,27,29,31,35-38,46,54,58,64,68,74,76,82,84, 87-89,98,99,101,103-106
Neural network	24 (11%)	5 (2%)	29 (14%)	16-18,20,24,29,30,36,46,54,60,68,69,73-76,78,82,84, 86-88,95,97,99,101,103,105
Regularization	14 (7%)	1 (0%)	15 (7%)	17,23,36,42,43,45,50,54,67,83,85,91,93,98,107
K-nearest neighbors	5 (2%)	1 (0%)	6 (3%)	20,37,46,58,87,106
Bayesian network	5 (2%)	0 (0%)	5 (2%)	20,22,27,32,87
Bayesian classifier	1 (0%)	1 (0%)	2 (1%)	46,106
Others[†]	12 (6%)	1 (0%)	13 (6%)	20,21,25,31,37,39,41,48,66,71,77,91,107
Super learner	2 (1%)	0 (0%)	2 (1%)	26,58
All	187 (89%)	23 (11%)	210 (100%)	

Note: Values are presented as n (%).
ML indicates machine learning; SES, simple exponential smoothing.
*Total number of ML models developed in 92 included studies.
[†]Others include C5.0 classification models, autoregressive integrated moving average model, SES model, partially linear additive quantile regression, hierarchical agglomerative clustering, and N-grams.

Among the rest of the studies with detailed descriptions about the model testing, the median number of observations in the testing data set was 777 (interquartile range 186-8438).

ML Methods and Performance Metrics Used

We identified a total of 210 different ML models and 236 model metrics across 92 included studies. The most common type of ML models was tree-based models such as random forest and gradient boosting (65 of 210, 31%), followed by support vector machine (30 of 210, 14%) and neural networks (29 of 210, 14%) (Table 3[16-107]). Our descriptive analysis showed that tree-based models were more likely to be used than traditional methods for clinical outcomes (eg, clinical event or disease incidence and treatment outcomes) than economic outcomes (eg, costs and healthcare utilization). There was no specific pattern in terms of the relationship between the type of ML methods and the type of data (See Figure S2 in Appendix 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.07.011).

In terms of the types of model outcome variable (ie, binary, categorical, or continuous), most studies used only one type, but there were 4 studies that used 2 different types, and hence, the total number was 96 types across 92 studies. Of those 96, the majority (77 of 96, 80%) were binary outcome and 19% (18 of 96) were continuous. There was only 1 study (1 of 92, 1%) that used categorical outcome.

Among all model performance metrics, studies have used area under the receiver operating characteristic most frequently (58 of 236, 25%), followed by sensitivity or recall (38 of 236, 16%) and specificity (28 of 236, 12%) (Table 4).

Discussion

We reviewed the recent trend in the applications of ML in HEOR with nonwearable data by conducting a scoping review of the published literature. Our review showed that the purpose of ML applications in HEOR has leaned toward forecasting rather than monitoring. We also found that the applications of ML have been mostly focused on clinical outcomes such as disease incidence and treatment outcomes rather than economic outcomes such as healthcare resource utilization and outcomes. The major types of data used for ML were EMR and primary data, and there has been relatively limited use of administrative claims data.

Despite the enthusiasm around the diverse roles that ML can potentially play in HEOR, the applications were mainly focused on predicting clinical outcomes, mostly using EMR data, with the aim of supporting providers' clinical practice. The underutilization of ML for economic outcomes and claims data could be simply because researchers are less interested in predicting economic outcomes than clinical outcomes. Although economic outcomes can be used as a proxy for patient health needs or health outcomes, they are mostly used in relatively limited settings such as coverage or pricing decision makings or evaluating the economic impact of health services or policies. In addition, claims data usually have a smaller dimensionality than other types of data, such as EMR data, and thus, the use of ML techniques is less likely to be relevant. Furthermore, few researchers would have access to claims data linked to other data with a higher dimensionality, which is also a reason why claims data are not frequently used with ML. In contrast, for EMR data, ML applications were relatively diverse compared with claims data, with a primary focus on predicting disease incidence or progression and less focus on economic outcomes such as healthcare resource utilization. Relatively active use of ML with EMR data may be due to a higher dimensionality, richness, and availability of EMR data and higher interest in predicting clinical outcomes. We believe that we can achieve broader applications of ML by linking different data sets, making it more relevant for ML use, and increasing the breadth of information available in the data. For example, claims data that have a relatively broad range of economic information, if combined with data with additional clinical and demographic

Table 4. Performance metrics used for ML models.

Performance metrics	Definition	n (%)
AUROC	Area under the sensitivity and (1-specificity) curve	58 (25)
Sensitivity or recall	TP / (TP + FN)	38 (16)
Specificity	TN / (TN + FP)	28 (12)
Precision or PPV	TP / (TP + FP)	25 (11)
Accuracy	(TP + TN) / (TP + FP + FN + TN)	23 (10)
MAE, MSE	The mean absolute error or the average of the squared difference between the original and predicted values	15 (6)
F1 score	2 × (Recall × Precision) / (Recall + Precision)	13 (6)
NPV	TN / (TN + FN)	8 (3)
Brier score	The average of the squared difference between the probability forecast and the actual outcome	5 (2)
Others*	N/A	19 (8)
Not reported	N/A	4 (2)
Total		236 (100)

AIC indicates Akaike information criterion; AUROC, area under the receiver operating characteristic; FN, false negatives; FP, false positives; ML, machine learning; MAE, mean absolute error; MSE, mean squared error; N/A, not applicable; TN, true negatives; TP, true positives.
*Other types of performance metrics include r-squared, log-loss, calibration intercept and slope, correlation coefficients, AIC, Rand index, mean average precision, weighted absolute percentage error, MSE, and MAE.

information, may be more suitable for ML use in studies that need a breadth of economic and clinical data.[108] Strategies for comprehensive data systems have been suggested by several studies, such as strengthening the existing data infrastructure and collaboration or partnership between public and private organizations.[108-110] In addition, there are practical guides for conducting linkages along with recommendations for good practice guidelines available for researchers.[109-111] Moreover, several investigators and government agencies have made efforts to link data sets to provide more comprehensive information about the patient population, such as the Surveillance, Epidemiology, and End Results-Medicare, Healthcare Cost and Utilization Project, National Cancer Institute Community Cancer Centers Program, and Health and Retirement Study.[110] Researchers could also consider linking traditional health data with nontraditional data, such as wearable data, which is unique regarding the types and timing of data collected. We found no study that linked nonwearable with wearable data in this review. Continued efforts to link data across different data types will broaden the roles of ML in HEOR. Continued efforts to link different data types will broaden the roles of ML in HEOR.

We also found that the use of ML in HEOR has been relatively limited in the clinical trial setting, compared with the real-world setting, accounting for only 5% of all included studies. We cannot overinterpret this finding because it could simply be because there are fewer clinical trials than observational studies and clinical trial data are less widely available than the real-world data. That said, it should be noted that the discussion of ML has been mostly focused on big observational data collected from the real world although there are potential roles of ML in clinical research as well.[112,113] For example, ML may help select target patient populations for clinical studies so that the required sample size can be reduced and also predict participant retention, improving the efficiency of patient recruitment and management. In addition, ML can allow researchers to track and monitor participants for the potential risk of adverse drug effects, thereby improving the safety of clinical trials. Moreover, post hoc analysis of trial data can be used with ML to identify treatment effects or to develop risk prediction models. ML may perform better than traditional modeling approaches in prediction. In addition, some ML techniques overcome the limitations of traditional subgroup analyses (eg, multiple comparisons and estimation bias) and permit detection of treatment heterogeneity when traditional subgroup analysis is well suited.[114] We encourage HEOR researchers to seek novel applications of ML by exploring how it can be leveraged in clinical trial settings.[112,113]

Consistent with part 1 of this study that was specific to wearable data, tree-based models were the most commonly used ML method in HEOR, followed by support vector machine, neural networks, and regularization. Nonsupervised learning has been rarely used, meaning that most of the studies aimed to map inputs to outputs. It should be noted that the most frequently used models in the included studies are not necessarily the best-fitting models. Some tree-based models, for example, might have been frequently used due to their ease of application. In addition, the process of building the trees is intuitive, making them easier for communication than other so-called black-box models. Although there might be additional costs to develop models that were not frequently encountered in our review, they could be explored if one is willing to pay for the potential incremental gains in the model performance. Given that no single type of ML model outperforms the others, researchers may need to explore several different types of models to find the one that maps the inputs and outputs well and has the appropriate level of human understandability.

Compared with the part 1 of our study, we see differences between the use of ML with wearable and nonwearable data in a few aspects. The most distinct difference was the purpose of the ML application—not surprisingly, ML has been mostly used for monitoring purposes with wearable data versus forecasting purposes with nonwearable data. The difference is likely due to the difference in the way of collecting the data—wearable devices measure and collect the outcomes that patients experience in their daily lives. In contrast, nonwearable data are more focused on the outcomes at the time of patient visits the clinic. Furthermore, studies using wearable data were less likely to be specific to certain disease areas, consistent with the fact that model outcomes being studied with wearable data were more likely to be measuring general health status or physical activities than disease-specific or treatment-specific outcomes. This finding is not surprising considering wearable data are relatively limited in the breadth of the information collected compared with EMR or registry data that often contain results from disease-specific tests, examinations, and interventions. Our study findings are consistent with an existing review study examining the use of ML in health-related research at a higher level.[12] The study looked at ML applications in population health contexts and showed EMR and investigator-generated data are the most commonly used data sources. In addition, they found tree-based models accounted for the largest proportion among all types of ML methods.

This study is not without limitations. There may be some relevant studies that we were unable to identify. The 2 key concepts of this review, ML and HEOR, are very broad, which made it challenging to find a search strategy that can comprehensively capture the eligible studies. Although focusing on 2020 to 2021 helps look at the trend of ML use in the most recent time period, the trend of ML use might have been affected by the pandemic. We expect that there might have been more studies on infectious diseases than usual because of the pandemic and more studies with predicting purposes, considering predicting the incidence or severity of COVID-19 has been of the major interest in this timeframe. We chose a 20% of random sample among the identified studies from PubMed for the same reasons. Although random sampling is not traditionally performed in conventional review studies, we believe that the randomization prevents the study results and conclusions from being affected by the sampling bias. Nevertheless, we acknowledge that our conclusions based on categories with a relatively small sample size may have a limited significance. With the small sample size, our results should be interpreted as a descriptive rather than as testing any specific hypothesis. Furthermore, our study only included studies written in English. Therefore, our results may not generalize to HEOR studies written in other languages. In addition, our study excluded studies that use omics and image data because they are not data types commonly used in HEOR. Nevertheless, we acknowledge that these data could be used in HEOR, and excluding these studies may bias the results—had they been included, we expect the proportion of studies examining clinical outcomes might have been greater.

Conclusion

Our study is the first scoping review specifically focused on the application of ML in HEOR. Although ML has been used in HEOR, its application has been predominantly for developing prediction models for clinical outcomes that can aid treatment decision making among providers. Our review shows that there remain opportunities for broader use of ML in HEOR, including predicting economic outcomes by leveraging claims data to help inform the examination of the cost-effectiveness of interventions. Although our study provides an overview of ML use in HEOR, future studies focusing on more specific research topics in HEOR could provide more detailed information about ML use.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2022.07.011.

Article and Author Information

Accepted for Publication: July 10, 2022

Published Online: August 18, 2022

doi: https://doi.org/10.1016/j.jval.2022.07.011

Author Affiliations: The Comparative Health Outcomes, Policy, and Economics (CHOICE) Institute, School of Pharmacy, University of Washington, Seattle, WA, USA (Lee, Schwartz, Bansal, Khor, Basu, Devine); Department of Health Services Research, Management & Policy, University of Florida, Gainesville, FL, USA (Hammarlund).

Correspondence: Woojung Lee, PharmD, Department of Pharmacy, The CHOICE Institute, University of Washington, Box 357630, Seattle, WA 98195-7630, USA. Email: wwlwj08@uw.edu

Author Contributions: *Concept and design:* Lee, Schwartz, Bansal, Khor, Hammarlund, Basu, Devine
Acquisition of data: Lee, Schwartz
Analysis and interpretation of data: Lee, Schwartz, Bansal, Khor, Basu, Devine
Drafting of the manuscript: Lee
Critical revision of the paper for important intellectual content: Lee, Schwartz, Bansal, Khor, Hammarlund, Basu, Devine
Statistical analysis: Lee, Bansal
Administrative, technical, or logistic support: Schwartz, Khor, Hammarlund
Supervision: Lee, Bansal, Devine

Conflict of Interest Disclosures: Drs Basu and Devine are editors for *Value in Health* and had no role in the peer-review process of this article. The other authors reported no additional conflicts of interest. No other disclosures were reported.

Funding/Support: This work was supported by the University of Washington School of Pharmacy corporate Advisory Board Health Tech Fund.

Role of the Funder/Sponsor: The funder participated in the study conceptualization and design and manuscript preparation and review. The funder had no role in the conduct of the study and collection, management, analysis, or approval of the manuscript.

Acknowledgment: The authors thank Diana N. Louden, a life sciences librarian at the University of Washington, for helping us develop search strategies for our scoping reviews.

REFERENCES

1. Kamel Boulos MN, Peng G, VoPham T. An overview of GeoAI applications in health and healthcare. *Int J Health Geogr*. 2019;18(1):7.
2. Zemouri R, Zerhouni N, Racoceanu D. Deep learning in the biomedical applications: recent and future status. *Appl Sci*. 2019;9(8):1526.
3. Wang H, Ma C, Zhou L. A brief review of machine learning and its application. In Proceedings from the 2009 International Conference on Information Engineering and Computer Science; Wuhan, China; December 19-20, 2009. 1-4.
4. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319(13):1317–1318.
5. Catalyst N. Healthcare big data and the promise of value-based care. *NEJM Catal*. 2018;4.
6. Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. *Health Aff (Millwood)*. 2014;33(7):1178–1186.
7. Etheredge LM. Rapid learning: a breakthrough agenda [published correction appears in *Health Aff (Millwood)*. 2014;33(9):1703]. *Health Aff (Millwood)*. 2014;33(7):1155–1162.
8. Real-world data: assessing electronic health records and medical claims data to support regulatory decision making for drug and biological products. US Food and Drug Administration. https://www.fda.gov/media/152503/download. Accessed June 7, 2022.
9. Sushmita S, Newman S, Marquardt J, et al. Population cost prediction on public healthcare data sets. In Proceedings of the 5th International Conference on Digital Health 2015; New York, NY; May 2015. 87-94.
10. Künzel SR, Sekhon JS, Bickel PJ, Yu B. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc Natl Acad Sci USA*. 2019;116(10):4156–4165.
11. Athey S, Imbens GW. Machine learning methods for estimating heterogeneous causal effects; 2015, Stanford Business School. https://www.gsb.stanford.edu/faculty-research/working-papers/machine-learning-estimating-heretogeneous-casual-effects. Accessed June 7, 2022.
12. Morgenstern JD, Buajitti E, O'Neill M, et al. Predicting population health with machine learning: a scoping review. *BMJ Open*. 2020;10(10):e037860.
13. Doupe P, Faghmous J, Basu S. Machine learning for health services researchers. *Value Health*. 2019;22(7):808–815.

14. About HEOR. The Professional Society for Health Economics and Outcomes Research. https://www.ispor.org/heor-resources/about-heor. Accessed January 11, 2022.
15. Lee W, Dayer V, Jiao B, Carlson JJ, Devine B, Veenstra DL. Use of real-world evidence in economic assessments of pharmaceuticals in the United States. *J Manag Care Spec Pharm*. 2021;27(1):5–14.
16. Abujaber A, Fadlalla A, Gammoh D, Abdelrahman H, Mollazehi M, El-Menyar A. Using trauma registry data to predict prolonged mechanical ventilation in patients with traumatic brain injury: machine learning approach. *PLoS One*. 2020;15(7):e0235231.
17. Zhang Y, Wang S, Hermann A, Joly R, Pathak J. Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women. *J Affect Disord*. 2021;279:1–8.
18. Zhou W, Wang Y, Gu X, et al. Importance of general adiposity, visceral adiposity and vital signs in predicting blood biomarkers using machine learning. *Int J Clin Pract*. 2021;75(1):e13664.
19. Zarkowsky DS, Nejim B, Hubara I, Hicks CW, Goodney PP, Malas MB. Deep learning and multivariable models select EVAR patients for short-stay discharge. *Vasc Endovasc Surg*. 2021;55(1):18–25.
20. Wu X-W, Yang H-B, Yuan R, Long E-W, Tong R-S. Predictive models of medication non-adherence risks of patients with T2D based on multiple machine learning algorithms. *BMJ Open Diabetes Res Care*. 2020;8(1):e001055.
21. Zhang X, Yu Y, Xiong F, Luo L. Prediction of daily blood sampling room visits based on ARIMA and SES Model. *Comput Math Methods Med*. 2020;2020:1720134.
22. Zhang Y, Huang L, Liu Y, Chen Q, Li X, Hu J. Prediction of mortality at one year after surgery for pertrochanteric fracture in the elderly via a Bayesian belief network. *Injury*. 2020;51(2):407–413.
23. Zhang Y, Zhu S, Yuan Z, et al. Risk factors and socio-economic burden in pancreatic ductal adenocarcinoma operation: a machine learning based analysis. *BMC Cancer*. 2020;20(1):1161.
24. Shashikumar SP, Wardi G, Paul P, et al. Development and prospective validation of a deep learning algorithm for predicting need for Mechanical Ventilation. *Chest*. 2021;159(6):2264–2273.
25. Pourghasemi HR, Pouyan S, Farajzadeh Z, et al. Assessment of the outbreak risk, mapping and infection behavior of COVID-19: application of the autoregressive integrated-moving average (ARIMA) and polynomial models. *PLoS One*. 2020;15(7):e0236238.
26. Shah AA, Devana SK, Lee C, Kianian R, van der Schaar M, SooHoo NF. Development of a novel, potentially universal machine learning algorithm for prediction of complications after total hip arthroplasty. *J Arthroplasty*. 2021;36(5):1655–1662.e1.
27. Roy S, Ghosh P. Factors affecting COVID-19 infected and death rates inform lockdown-related policymaking. *PLoS One*. 2020;15(10):e0241165.
28. Schwartz B, Cohen ZD, Rubel JA, Zimmermann D, Wittmann WW, Lutz W. Personalized treatment selection in routine care: integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. *Psychother Res*. 2021;31(1):33–51.
29. Savarraj JP, Hergenroeder GW, Zhu L, et al. Machine learning to predict delayed cerebral ischemia and outcomes in subarachnoid hemorrhage. *Neurology*. 2021;96(4):e553–e562.
30. Srinivas S. A machine learning-based approach for predicting patient punctuality in ambulatory care centers. *Int J Environ Res Public Health*. 2020;17(10):3703.
31. Pennell C, Polet C, Arthur LG, Grewal H, Aronoff S. Risk assessment for intra-abdominal injury following blunt trauma in children: derivation and validation of a machine learning model. *J Trauma Acute Care Surg*. 2020;89(1):153–159.
32. Perkins ZB, Yet B, Sharrock A, et al. Predicting the outcome of limb revascularization in patients with lower-extremity arterial trauma: development and external validation of a supervised machine-learning algorithm to support surgical decisions. *Ann Surg*. 2020;272(4):564–572.
33. Peng L-N, Hsiao F-Y, Lee W-J, Huang S-T, Chen L-K. Comparisons between hypothesis- and data-driven approaches for multimorbidity frailty index: a machine learning approach. *J Med Internet Res*. 2020;22(6):e16213.
34. Tsiklidis EJ, Sims C, Sinno T, Diamond SL. Using the National Trauma Data Bank (NTDB) and machine learning to predict trauma patient mortality at admission. *PLoS One*. 2020;15(11):e0242166.
35. Küpper C, Stroth S, Wolff N, et al. Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. *Sci Rep*. 2020;10(1):4805.
36. Kunze KN, Polce EM, Nwachukwu BU, Chahla J, Nho SJ. Development and internal validation of supervised machine learning algorithms for predicting clinically significant functional improvement in a mixed population of primary hip arthroscopy. *Arthroscopy*. 2021;37(5):1488–1497.
37. Kanda E, Epureanu BI, Adachi T, et al. Application of explainable ensemble artificial intelligence model to categorization of hemodialysis-patient and treatment using nationwide-real-world data in Japan. *PLoS One*. 2020;15(5):e0233491.
38. Hilbert K, Jacobi T, Kunas SL, et al. Identifying CBT non-response among OCD outpatients: a machine-learning approach. *Psychother Res*. 2021;31(1):52–62.
39. Muscas G, Matteuzzi T, Becattini E, et al. Development of machine learning models to prognosticate chronic shunt-dependent hydrocephalus after aneurysmal subarachnoid hemorrhage. *Acta Neurochir (Wien)*. 2020;162(12):3093–3105.
40. Muremyi R, Haughton D, Kabano I, Niragire F. Prediction of out-of-pocket health expenditures in Rwanda using machine learning techniques. *Pan Afr Med J*. 2020;37:357.
41. Mazumdar M, Lin J-YJ, Zhang W, et al. Comparison of statistical and machine learning models for healthcare cost data: a simulation study motivated by oncology care model (OCM) data. *BMC Health Serv Res*. 2020;20(1):350.
42. Kochav SM, Raita Y, Fifer MA, et al. Predicting the development of adverse cardiac events in patients with hypertrophic cardiomyopathy using machine learning. *Int J Cardiol*. 2021;327:117–124.
43. Ong MS, Klann JG, Lin KJ, et al. Claims-based algorithms for identifying patients with pulmonary hypertension: a comparison of decision rules and machine-learning approaches. *J Am Heart Assoc*. 2020;9(19):e016648.
44. Moro A, Mehta R, Tsilimigras DI, et al. Prognostic factors differ according to KRAS mutational status: a classification and regression tree model to define prognostic groups after hepatectomy for colorectal liver metastasis. *Surgery*. 2020;168(3):497–503.
45. Marić I, Tsur A, Aghaeepour N, et al. Early prediction of preeclampsia via machine learning. *Am J Obstet Gynecol MFM*. 2020;2(2):100100.
46. Lötsch J, Alfredsson L, Lampa J. Machine-learning–based knowledge discovery in rheumatoid arthritis–related registry data to identify predictors of persistent pain. *Pain*. 2020;161(1):114–126.
47. Martini ML, Neifert SN, Oermann EK, et al. Machine learning with feature domains elucidates candidate drivers of hospital readmission following spine surgery in a large single-center patient cohort. *Neurosurgery*. 2020;87(4):E500–E510.
48. Lerner J, Ruppenkamp J, Etter K, et al. Preoperative behavioral health, opioid, and antidepressant utilization and 2-year costs after spinal fusion—revelations from cluster analysis. *Spine (Phila Pa 1976)*. 2020;45(2):E90–E98.
49. Kamal S, Urata J, Cavassini M, et al. Random forest machine learning algorithm predicts virologic outcomes among HIV infected adults in Lausanne, Switzerland using electronically monitored combined antiretroviral treatment adherence. *AIDS Care*. 2021;33(4):530–536.
50. Jain D, Durand W, Burch S, Daniels A, Berven S. Machine learning for predictive modeling of 90-day readmission, major medical complication, and discharge to a facility in patients undergoing long segment posterior lumbar spine fusion. *Spine (Phila Pa 1976)*. 2020;45(16):1151–1160.
51. Iwamoto Y, Imura T, Tanaka R, et al. Development and validation of machine learning-based prediction for dependence in the activities of daily living after stroke inpatient rehabilitation: a decision-tree analysis. *J Stroke Cerebrovasc Dis*. 2020;29(12):105332.
52. Lipschuetz M, Guedalia J, Rottenstreich A, et al. Prediction of vaginal birth after cesarean deliveries using machine learning. *Am J Obstet Gynecol*. 2020;222(6):613.e1–613.e12.
53. Kumar V, Roche C, Overman S, et al. Using machine learning to predict clinical outcomes after shoulder arthroplasty with a minimal feature set. *J Shoulder Elbow Surg*. 2021;30(5):e225–e236.
54. Kunze KN, Polce EM, Rasio J, Nho SJ. Machine learning algorithms predict clinically significant improvements in satisfaction after hip arthroscopy. *Arthroscopy*. 2021;37(4):1143–1151.
55. Merath K, Hyer J, Mehta R, et al. Use of machine learning for prediction of patient risk of postoperative complications after liver, pancreatic, and colorectal surgery. *J Gastrointest Surg*. 2020;24(8):1843–1851.
56. Tu W, Chen PA, Koenig N, et al. Machine learning models reveal neurocognitive impairment type and prevalence are associated with distinct variables in HIV/AIDS. *J Neurovirol*. 2020;26(1):41–51.
57. Sinha P, Churpek MM, Calfee CS. Machine learning classifier models can identify acute respiratory distress syndrome phenotypes using readily available clinical data. *Am J Respir Crit Care Med*. 2020;202(7):996–1004.
58. Shtar G, Rokach L, Shapira B, Nissan R, Hershkovitz A. Using machine learning to predict rehabilitation outcomes in postacute hip fracture patients. *Arch Phys Med Rehabil*. 2021;102(3):386–394.
59. Muñoz B, Schobel SA, Lisboa FA, et al. Clinical risk factors and inflammatory biomarkers of post-traumatic acute kidney injury in combat patients. *Surgery*. 2020;168(4):662–670.
60. Mollalo A, Rivera KM, Vahedi B. Artificial neural network modeling of novel coronavirus (COVID-19) incidence rates across the continental United States. *Int J Environ Res Public Health*. 2020;17(12):4204.
61. Morel D, Kalvin CY, Liu-Ferrara A, Caceres-Suriel AJ, Kurtz SG, Tabak YP. Predicting hospital readmission in patients with mental or substance use disorders: a machine learning approach. *Int J Med Inform*. 2020;139:104136.
62. Menardi A, Bertagnoni G, Sartori G, Pastore M, Mondini S. Past life experiences and neurological recovery: the role of cognitive reserve in the rehabilitation of severe post-anoxic encephalopathy and traumatic brain injury. *J Int Neuropsychol Soc*. 2020;26(4):394–406.
63. Hu L, Liu B, Ji J, Li Y. Tree-based machine learning to identify and understand major determinants for stroke at the neighborhood level. *J Am Heart Assoc*. 2020;9(22):e016745.
64. Hopkins SC, Ogirala A, Loebel A, Koblan KS. Characterization of specific and distinct patient types in clinical trials of acute schizophrenia using an uncorrelated PANSS score matrix transform (UPSM). *Psychiatry Res*. 2020;294:113569.

65. Heo J, Park SJ, Kang S-H, Oh CW, Bang JS, Kim T. Prediction of intracranial aneurysm risk using machine learning. *Sci Rep*. 2020;10(1):6921.
66. Wu W, Zhang Y, Jiang J, et al. An electroencephalographic signature predicts antidepressant response in major depression. *Nat Biotechnol*. 2020;38(4):439–447.
67. Wong ES, Schuttner L, Reddy A. Does machine learning improve prediction of VA primary care reliance? *Am J Manag Care*. 2020;26(1):40–44.
68. Wong ZS-Y, So H, Kwok BS, Lai MW, Sun DT. Medication-rights detection using incident reports: a natural language processing and deep neural network approach. *Health Inform J*. 2020;26(3):1777–1794.
69. Sherazi SWA, Jeong YJ, Jae MH, Bae JW, Lee JY. A machine learning–based 1-year mortality prediction model after hospital discharge for clinical patients with acute coronary syndrome. *Health Inform J*. 2020;26(2):1289–1304.
70. Song X, Yu AS, Kellum JA, et al. Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. *Nat Commun*. 2020;11(1):5668.
71. Sabbah MS, Fayyaz AU, De Denus S, et al. Obese-inflammatory phenotypes in heart failure with preserved ejection fraction. *Circ Heart Fail*. 2020;13(8):e006414.
72. Mattie H, Reidy P, Bachtiger P, et al. A framework for predicting impactability of digital care management using machine learning methods. *Popul Health Manag*. 2020;23(4):319–325.
73. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, Kalhori SRN. Predicting COVID-19 incidence through analysis of google trends data in Iran: data mining and deep learning pilot study. *JMIR Public Health Surveill*. 2020;6(2):e18828.
74. Lee S, Eun Y, Kim H, Cha HS, Koh EM, Lee J. Machine learning to predict early TNF inhibitor users in patients with ankylosing spondylitis. *Sci Rep*. 2020;10(1):20299.
75. Lee HC, Park JS, Choe JC, et al. Prediction of 1-year mortality from acute myocardial infarction using machine learning. *Am J Cardiol*. 2020;133:23–31.
76. Hamilton EF, Dyachenko A, Ciampi A, Maurel K, Warrick PA, Garite TJ. Estimating risk of severe neonatal morbidity in preterm births under 32 weeks of gestation. *J Matern Fetal Neonatal Med*. 2020;33(1):73–80.
77. Harris AH, Kuo AC, Bowe TR, Manfredi L, Lalani NF, Giori NJ. Can machine learning methods produce accurate and easy-to-use preoperative prediction models of one-year improvements in pain and functioning after knee arthroplasty? *J Arthroplasty*. 2021;36(1):112–117.e6.
78. Ng SHX, Rahman N, Ang IYH, et al. Characterising and predicting persistent high-cost utilisers in healthcare: a retrospective cohort study in Singapore. *BMJ Open*. 2020;10(1):e031622.
79. Prediction of a due date based on the pregnancy history data using machine learning. In: Blobel B, ed. *pHealth 2020 Proceedings of the 17th International Conference on Wearable Micro and Nano Technologies for Personalized Health*. 2020.
80. Churpek MM, Carey KA, Edelson DP, et al. Internal and external validation of a machine learning risk score for acute kidney injury. *JAMA Netw Open*. 2020;3(8):e2012892.
81. Chocron A, Efraim R, Mandel F, et al. Machine learning for nocturnal mass diagnosis of atrial fibrillation in a population at risk of sleep-disordered breathing. *Physiol Meas*. 2020;41(10):104001.
82. Arvind V, London DA, Cirino C, Keswani A, Cagle PJ. Comparison of machine learning techniques to predict unplanned readmission following total shoulder arthroplasty. *J Shoulder Elbow Surg*. 2021;30(2):e50–e59.
83. Escobar GJ, Soltesz L, Schuler A, Niki H, Malenica I, Lee C. Prediction of obstetrical and fetal complications using automated electronic health record data. *Am J Obstet Gynecol*. 2021;224(2):137–147.e7.
84. Fernandes MPB, de la Hoz MA, Rangasamy V, Subramaniam B. Machine learning models with preoperative risk factors and intraoperative hypotension parameters predict mortality after cardiac surgery. *J Cardiothorac Vasc Anesth*. 2021;35(3):857–865.
85. Betts KS, Kisely S, Alati R. Predicting postpartum psychiatric admission using a machine learning approach. *J Psychiatr Res*. 2020;130:35–40.
86. Arbour KC, Luu AT, Luo J, et al. Deep learning to estimate RECIST in patients with NSCLC treated with PD-1 blockade. *Cancer Discov*. 2021;11(1):59–67.
87. Dovgan E, Gradišek A, Luštrek M, et al. Using machine learning models to predict the initiation of renal replacement therapy among chronic kidney disease patients. *PLoS One*. 2020;15(6):e0233976.
88. Choi J, Kwon L-N, Lim H, Chun H-W. Gender-based analysis of risk factors for dementia using senior cohort. *Int J Environ Res Public Health*. 2020;17(19):7274.
89. Glauser T, Santel D, DelBello M, et al. Identifying epilepsy psychiatric comorbidities with machine learning. *Acta Neurol Scand*. 2020;141(5):388–396.
90. Cary Jr MP, Zhuang F, Draelos RL, et al. Machine learning algorithms to predict mortality and allocate palliative care for older patients with hip fracture. *J Am Med Dir Assoc*. 2021;22(2):291–296.
91. Chou C-A, Cao Q, Weng S-J, Tsai C-H. Mixed-integer optimization approach to learning association rules for unplanned ICU transfer. *Artif Intell Med*. 2020;103:101806.
92. Behravan H, Hartikainen JM, Tengström M, Kosma VM, Mannermaa A. Predicting breast cancer risk using interacting genetic and demographic factors and machine learning. *Sci Rep*. 2020;10(1):11044.
93. Ajana S, Cougnard-Grégoire A, Colijn JM, et al. Predicting progression to advanced age-related macular degeneration from clinical, genetic, and lifestyle factors using machine learning. *Ophthalmology*. 2021;128(4):587–597.
94. Gómez-Ramírez J, Ávila-Villanueva M, Fernández-Blázquez MÁ. Selecting the most important self-assessed features for predicting conversion to mild cognitive impairment with random forest and permutation-based methods. *Sci Rep*. 2020;10(1):20630.
95. Yang C, Ojha B, Aranoff ND, et al. Classification of aortic stenosis before and after transcatheter aortic valve replacement using cardio-mechanical modalities. In 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Montreal, QC; July 20-24, 2020. 2820-2823.
96. Muñoz-Almaraz FJ, Climent MT, Guerrero MD, Moreno L, Pardo J. A machine learning approach to design an efficient selective screening of mild cognitive impairment. *J Vis Exp*. 2020. https://doi.org/10.3791/59649.
97. Liu K, Xia W, Qiang M, et al. Deep learning pathological microscopic features in endemic nasopharyngeal cancer: prognostic value and protentional role for individual induction chemotherapy. *Cancer Med*. 2020;9(4):1298–1306.
98. Lin W-C, Goldstein IH, Hribar MR, Sanders DS, Chiang MF. Predicting wait times in pediatric ophthalmology outpatient clinic using machine learning. *AMIA Annu Symp Proc*. 2020;2019:1121–1128.
99. Lin AX, Ho AFW, Cheong KH, et al. Leveraging machine learning techniques and engineering of multi-nature features for national daily regional ambulance demand prediction. *Int J Environ Res Public Health*. 2020;17(11):4179.
100. Li Y, Chen M, Lv H, Yin P, Zhang L, Tang P. A novel machine-learning algorithm for predicting mortality risk after hip fracture surgery. *Injury*. 2021;52(6):1487–1493.
101. Lee S-Y, Chinnam RB, Dalkiran E, Krupp S, Nauss M. Prediction of emergency department patient disposition decision for proactive resource allocation for admission. *Health Care Manag Sci*. 2020;23(3):339–359.
102. Howard EP, Morris JN, Schachter E, Schwarzkopf R, Shepard N, Buchanan ER. Machine-learning modeling to predict hospital readmission following discharge to post-acute care. *J Am Med Dir Assoc*. 2021;22(5):1067–1072.e29.
103. Hilbert K, Kunas SL, Lueken U, Kathmann N, Fydrich T, Fehm L. Predicting cognitive behavioral therapy outcome in the outpatient sector based on clinical routine data: a machine learning approach. *Behav Res Ther*. 2020;124:103530.
104. Gupta U, Bansal H, Joshi D. An improved sex-specific and age-dependent classification model for Parkinson's diagnosis using handwriting measurement. *Comput Methods Programs Biomed*. 2020;189:105305.
105. Feller DJ, Bear Don't Walk Iv OJ, Zucker J, Yin MT, Gordon P, Elhadad N. Detecting social and behavioral determinants of health with structured and free-text clinical data. *Appl Clin Inform*. 2020;11(1):172–181.
106. Aljameel SS, Alabbad DA, Alzahrani NA, et al. A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent COVID-19 outbreaks in Saudi Arabia. *Int J Environ Res Public Health*. 2021;18(1):218.
107. Goldberg SB, Flemotomos N, Martinez VR, et al. Machine learning and natural language processing in psychotherapy research: alliance as example use case. *J Couns Psychol*. 2020;67(4):438.
108. Crown WH. Potential application of machine learning in health outcomes research and some statistical cautions. *Value Health*. 2015;18(2):137–140.
109. Dusetzina SB, Tyree S, Meyer A-M, Meyer A, Green L, Carpenter WR. *Linking Data for Health Services Research: A Framework and Instructional Guide*. Rockville, MD: Agency for Healthcare Research and Quality (US); 2014.
110. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health Services Research and data linkages: issues, methods, and directions for the future. *Health Serv Res*. 2010;45(5 Pt 2):1468–1488.
111. Harron K, Dibben C, Boyd J, et al. Challenges in administrative data linkage for research. *Big Data Soc*. 2017;4(2):2053951717745678.
112. Weissler EH, Naumann T, Andersson T, et al. The role of machine learning in clinical research: transforming the future of evidence generation [published correction appears in *Trials*. 2021;22(1):593]. *Trials*. 2021;22(1):537.
113. Shah P, Kendall F, Khozin S, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. *NPJ Digit Med*. 2019;2:69.
114. Rigdon J, Baiocchi M, Basu S. Preventing false discovery of heterogeneous treatment effect subgroups in randomized trials. *Trials*. 2018;19(1):382.