



OPEN ACCESS



Use of electronic medical records in development and validation of risk prediction models of hospital readmission: systematic review

Elham Mahmoudi,^{1,2} Neil Kamdar,^{2,3,4,5} Noa Kim,¹ Gabriella Gonzales,^{6,1} Karandeep Singh,^{7,10} Akbar K Waljee^{8,9,11}

For numbered affiliations see end of the article.

Correspondence to: E Mahmoudi Mahmoudi@med.umich.edu (or @Mahmoudi_E on Twitter; ORCID 0000-0002-9746-8165)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2020;369:m958

<http://dx.doi.org/10.1136/bmj.m958>

Accepted: 27 February 2020

ABSTRACT

OBJECTIVE

To provide focused evaluation of predictive modeling of electronic medical record (EMR) data to predict 30 day hospital readmission.

DESIGN

Systematic review.

DATA SOURCE

Ovid Medline, Ovid Embase, CINAHL, Web of Science, and Scopus from January 2015 to January 2019.

ELIGIBILITY CRITERIA FOR SELECTING STUDIES

All studies of predictive models for 28 day or 30 day hospital readmission that used EMR data.

OUTCOME MEASURES

Characteristics of included studies, methods of prediction, predictive features, and performance of predictive models.

RESULTS

Of 4442 citations reviewed, 41 studies met the inclusion criteria. Seventeen models predicted risk of readmission for all patients and 24 developed predictions for patient specific populations, with 13 of those being developed for patients with heart conditions. Except for two studies from the UK and Israel, all were from the US. The total sample size for each model ranged between 349 and 1 195 640. Twenty five models used a split sample validation technique. Seventeen of 41 studies reported C

statistics of 0.75 or greater. Fifteen models used calibration techniques to further refine the model. Using EMR data enabled final predictive models to use a wide variety of clinical measures such as laboratory results and vital signs; however, use of socioeconomic features or functional status was rare. Using natural language processing, three models were able to extract relevant psychosocial features, which substantially improved their predictions. Twenty six studies used logistic or Cox regression models, and the rest used machine learning methods. No statistically significant difference (difference 0.03, 95% confidence interval -0.0 to 0.07) was found between average C statistics of models developed using regression methods (0.71, 0.68 to 0.73) and machine learning (0.74, 0.71 to 0.77).

CONCLUSIONS

On average, prediction models using EMR data have better predictive performance than those using administrative data. However, this improvement remains modest. Most of the studies examined lacked inclusion of socioeconomic features, failed to calibrate the models, neglected to conduct rigorous diagnostic testing, and did not discuss clinical impact.

Introduction

Hospitals across the US continue to be under scrutiny to reduce their 30 day readmission rates (hereafter readmission), as a measure of both hospital quality and cost reduction. The Hospital Readmissions Reduction Program is a Medicare value based program that since October 2012 has started reducing payments to hospitals with excess readmissions.¹ Between 2007 and 2015, readmission rates for specific conditions dropped from 21.5% to 17.5%.² This has been largely attributed to investments by hospitals to enhance their discharge processes,² which include providing better medication reconciliation, educating patients and their care givers regarding continuity of care, and implementing follow-up processes for discharged patients. However, implementation of an effective discharge process in hospitals is time consuming and expensive. The development of readmission risk tools has increased sharply in recent years to enable precise identification of patients at high risk and inform a more efficient use of post-discharge care coordination. However, because of the complexity of inpatient care and discharge processes, achieving a high sensitivity and specificity in predicting who is at risk of readmission and why is still a work in progress.

WHAT IS ALREADY KNOWN ON THIS TOPIC

The development of tools to predict the risk of 30 day hospital readmission and thus enable identification of patients at high risk has increased sharply in recent years

However, achieving a high sensitivity and specificity in predicting who is at risk of readmission and why is still a work in progress

The accuracy and reliability of risk prediction models largely depend on predictors and methods of development, validation, calibration, and clinical utility

WHAT THIS STUDY ADDS

On average, risk prediction models using electronic medical records have better predictive performance than those using administrative data, but this improvement remains modest

The quality and integrity of electronic medical records are concerning and pose significant barriers to effectively leveraging these data to develop accurate and precise risk assessment tools

Most studies did not account for salient socioeconomic features, failed to calibrate their models, and lacked careful assessment of the clinical utilities and implementation of the developed tools

The accuracy and reliability of risk models largely depend on predictors and methods of development, validation, calibration, and clinical utility.³ In the context of choosing an appropriate set of predictors, administrative data are inherently limited, primarily due to the lack of clinical specificity for conditions and laboratory results. With recent multibillion dollar investments in electronic medical records (EMRs) and their increasing use and application in healthcare systems,⁴ the use of machine learning methods in medicine has also expanded. Thus, the past few years has seen a surge in the development of highly sophisticated predictive models using EMRs. Two previously published systematic reviews of predictive models of readmission—regardless of the data source used or whether the model was validated—assessed predictive models up to 2015.^{5–6} Gaps exist in the knowledge about predictive models of readmission that leverage the use of EMRs and new methods of prediction.

This study focuses on validated predictive models of readmission that specifically use EMR data. We adopted the systematic review guide for evaluation of prediction model performance.⁷ The objectives of this study were to evaluate the variation in predicting readmission for all patients versus patient specific populations, to examine the properties of the EMR based candidate features, to assess differences in performance between traditional regression and machine learning models, and to assess the quality of the studies.

Methods

Information sources and search

We searched Ovid Medline, Ovid Embase, CINAHL, Web of Science, and Scopus by using an inclusive combination of exploded MeSH subject headings, keywords, and title, abstract, and full text keywords, with and without adjacencies when available, with a publication date range of 1 January 2015 to 1 January 2019. The last electronic database search took place in April 2019. We imported all citations into electronic citation management software (EndNote X9). Supplementary tables A–C provide detailed information on inclusion and exclusion criteria and on our search strategy.

Eligibility criteria

Studies eligible for inclusion were peer reviewed and published between 1 January 2015 and 1 January 2019. We included only studies that developed and validated a predictive model of hospital readmission within 28 or 30 days after initial discharge. We excluded studies that did not use EMR data in the development or validation of the model, studies published before 2015 owing to overlap with previous reviews,^{5–6} studies not published in English, and conference abstract only references (supplementary table A). We did not do an extensive hand search for this systematic review.

Study selection

After de-duplication, two authors (EM and GG) screened our initial 3506 citations for title and abstract relevance. We excluded 3206 records and accessed 300 resulting citations in their full text form. Two authors evaluated each article independently by using the inclusion and exclusion criteria shown in figure 1. Discrepancies between reviewers were resolved through additional review during group discussions.

Data extraction

Two authors (EM and GG) extracted data from the final included studies to profile each model's population (table 1 and table 2), candidate features (supplementary table D), model description (supplementary table E), and quality assessment (supplementary table F). To ease the cross linkage between tables 1 and 2 and the supplements, we organized all supplementary tables similarly. Firstly, we separated the included studies into two general categories: all patient populations and specific patient populations. We then listed studies in each group alphabetically according to lead author's last name.

Data synthesis

The wide heterogeneity of the included models did not permit a quantitative meta-analysis of their performance; however, we provide a qualitative review and synthesis of population studied and model characteristics. To analyze the differences between studies that used machine learning methods and those that used traditional regression or between those for all patient populations and those developed for specific populations, we assumed that every study was weighted equally regardless of the number of patients and/or methods used. If more than one model was used in a study, we chose the C statistic representing the maximum for that study. We report validation C statistics in this review; however, when the study was ambiguous about the C statistic being from either the development or validation dataset, we assumed they were being reported from the validation cohort. Finally, on further analysis, we calculated 95% confidence intervals for the C statistics of different study groups. We also calculated the 95% confidence intervals for the difference in the mean C statistic between the two study groups to ascertain potentially significant differences in concordance.

Results

From 3506 titles and abstracts (after removing 937 duplicates), we selected 300 articles for complete text review. Our final set included 41 studies that met our inclusion criteria (fig 1). We divided these studies on the basis of their population cohort into all patient populations (n=17, including one intensive care unit and one emergency department readmission) and patient specific populations (n=24). Most patient specific models were for heart conditions (n=13).^{24–27 30 32 34 37 38 43 45–47} The remainder were

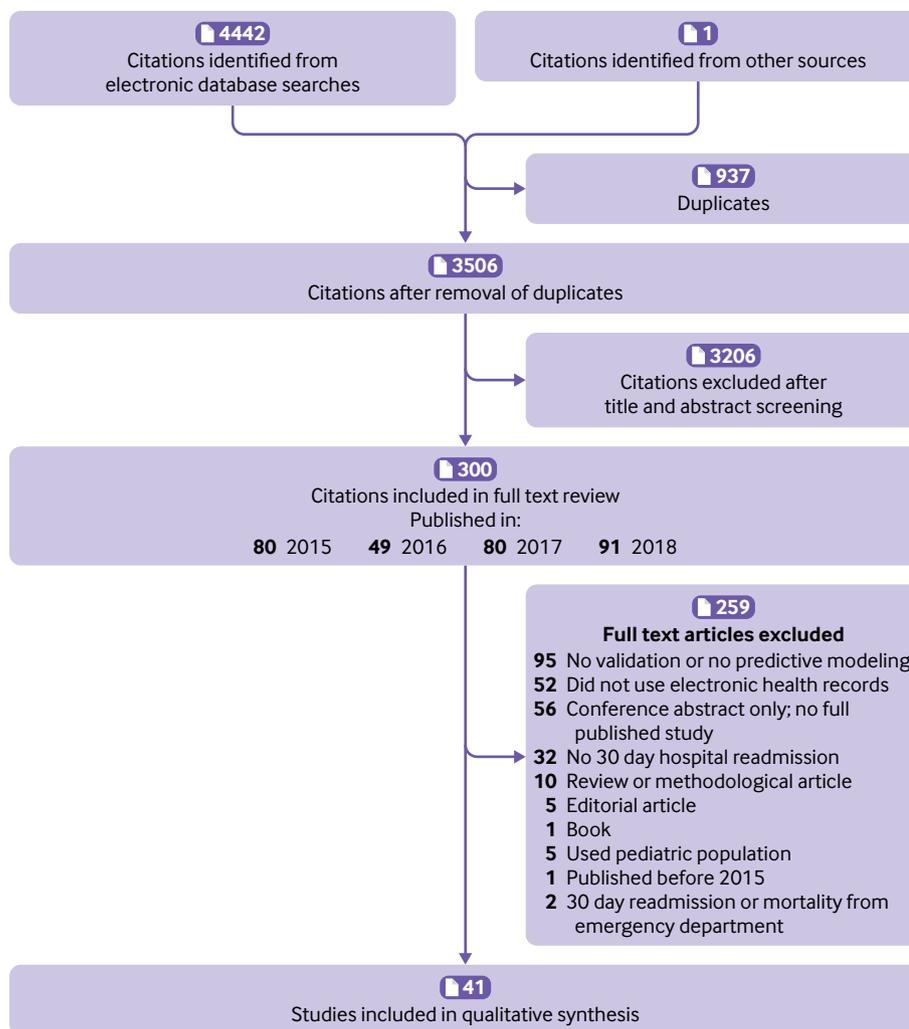


Fig 1 | Schematic flow diagram of selected studies[A: I believe the number of full text articles excluded should be 259 rather than 257]

based on readmission among patients with diabetes (n=4),^{28 33 40 41} kidney transplantation (1),⁴⁴ hemodialysis (1),²⁹ low back surgery (1),³⁶ pneumonia (2),^{31 35} lupus (1),³⁹ and psychiatric conditions (1).⁴² Thirty nine studies were based on data from US hospitals, and two were from other developed countries (the UK¹¹ and Israel²⁰).

The total sample size in each model ranged from 349 to 1 195 640.^{21 29} All validations were done internally; most were conducted through retrospective validation (n=37) and used split sample (n=24) or cross validation (n=11) methods. The C statistics ranged between 0.52 and 0.90,^{23 24} with 17 studies reporting a C statistic of 0.75 or greater.^{11 12 14-17 19 23 29 33 34 36 37 42 43 46 47}

Characteristics of patients and hospitals

Table 1 and table 2 show characteristics of patient populations and hospitals. Seventeen studies developed predictive models of readmission for all patient populations (table 1). Most studies included adults 18 years or older (n=9). Twelve studies used data from multiple hospitals. Included centers were non-academic (n=9),^{8 9 11-15 22 24} academic (n=4),^{10 16 19 23} or a

combination of both (n=3).^{18 20 21} Observed readmission rates for these sets of models were between 6% and 23%.^{9 24}

Twenty four studies developed predictive models of readmission among specific patient populations (table 2). Most (13/24 studies) of these models were developed for patients admitted with heart conditions.^{24-27 30 32 34 37 38 43 45-47} All patient specific studies included adults 18 years or older. Ten studies used data from multiple hospitals.^{25 26 28 30 32 34 35 37 39 47} Data came from academic centers (n=9),^{24 29 31 33 36 40 41 42 44} non-academic centers (n=3),^{27 38 43} or a combination of both (n=10).^{25 26 28 30 32 35 37 39 47} Observed readmission rates for patient specific readmissions ranged between 5.9% and 54%.^{36 38}

Candidate features and predictors

Supplementary table D summarizes the features used in the predictive models. We categorized the features into five groups: clinical data, demographics, healthcare encounter history, functional status, and socioeconomic status. Using EMR data, detailed clinical and healthcare encounter data such as admission

Table 1 | Characteristics of patients and hospitals in studies that included all patient populations

Study	Study population	Hospital type	Multicenter	Total sample size		Observed readmission rate (%)
				Derivation	Validation	
Amarasingham et al, 2015 ⁸	Adults 18+	Non-academic/ large community	Yes	19 831	19 773	12.7
Brindise et al, 2018 ⁹	All patients	Non-academic/ large community	Yes	8,814	4,407	23
Chen et al, 2016 ¹⁰	NA	Academic	No	15 629	1,897	8.3
Damery et al, 2017 ¹¹	18+	Non-academic/ large community	No	51 747	51 747	7.7
Escobar et al, 2015 ¹²	18+	Non-academic/ large community	Yes	179 978	180 058	Any: 14.5; non-elective: 12.5
Greenwood et al, 2018 ¹³	18+	Non-academic/ large community	Yes	39 155	NA	11.1%
Hao et al, 2015 ¹⁴	All patients	Non-academic/ large community	Yes	24 810	Retrospective: 24 857; prospective: 118 951	Retrospective: 13.2; prospective: 14.7
Jamei et al, 2017 ¹⁵	All patients	Non-academic/ large community	Yes	268 652	67 163	9.7
Logue et al, 2016 ¹⁶	18+	Academic	No	958	Bootstrap cross validation	14
Morris et al, 2016 ¹⁷	18+	VA surgical quality improvement	Yes	213 697	23 744	11.1
Nguyen et al, 2016 ¹⁸	All patients	Academic and non-academic	Yes	16 492	16 430	12.7
Rajkomar et al, 2018 ¹⁹	18+	Academic	Yes	194 470	21 751	Hospital A: 10.5; hospital B: 15.1
Shadmi et al, 2015 ²⁰	18+	Academic and non-academic	Yes	22 406	11 233	15.2
Tabak et al, 2017 ²¹	18+	Academic and non-academic	Yes	836 992	358 648	11.9
Tong et al, 2016 ²²	All patients	Non-academic/ large community	Yes	80 000	80 000*	11.5
Walsh et al, 2017 ²³	All patients	Academic	No	92 530	27 470	All cause: 13.4
Wang et al, 2018 ²⁴	NA	Non-academic/ large community	No	41 503/700	60:15:25 split for training, validation, and testing	Hospital data: 6; operating room data: 17.7

NA=not available; VA=Veterans Affairs.

*Different sample sizes of 2500, 5000, 20 000, and 80 000 for derivation and validation of four different models were considered. Results shown are for sample size of 80 000.

type and discharge location, primary and additional diagnoses, morbidities, laboratory results, vital signs, type and number of drugs, and basic demographics such as age, sex, race and ethnicity, and insurance type were readily available and thus examined in most of the predictive models. Additionally, Escobar et al and Morris et al used length of operating room stay in hours as a proxy for complexity of surgical procedure if the inpatient hospital stay included any surgical procedure.^{12 17} Being admitted to an intensive care unit and number of procedures during the index hospital stay have also been used as proxies for the complexity of a patient's condition.^{15 18 19 22 25 27 28 30 38 43}

A few studies used composite clinical scores that are not readily available to account for severity of conditions for patients admitted to hospital. For example, Tong et al used the Braden Score to indicate risk of pressure ulcers,²² and Escobar et al used the Comorbidity Point Score or COPS2,¹² which uses 45 of the 70 possible Hierarchical Condition Categories originally developed by the Centers for Medicare and Medicaid Services to measure the severity of a patient's comorbidity. Other noteworthy composite scores included severity of illness on the day of admission and discharge based on the Laboratory Acute Physiology Score,¹² Acute Laboratory Risk of Mortality Score,²¹ polypharmacy (more than six medicines),^{16 22} surgical complications,⁴¹ number of laboratory results marked as "high," "low," or

"abnormal,"⁹ use of specific drugs among patients admitted for heart failure,³⁸ or use of 10 or more drugs at the time of admission among hemodialysis patients.²⁹

Functional status is usually not recorded in structured EMR data. Only seven studies included measures of disability or limitations on activities of daily living in their models.^{13 16 17 20 34 36 38} Shadmi et al used a disability indicator, which is routinely collected in Clalit Health Services data in Israel,²⁰ and this proved to be a top predictor for readmission. Morris et al used functional status available via Veterans Affairs data and nurses' notes,¹⁷ and McGirt et al used questionnaire data in addition to EMRs to collect this information.³⁶

Sixteen studies considered various socioeconomic proxies for socioeconomic status as candidate predictors for readmission.^{9 13-18 20 21 28 30 33 35 38 40 41 44} These studies used nurses' notes, self-reported patient questionnaires, or census 2010 block level or zip code level aggregate data to include features such as income and education. Despite the cited importance of care giver availability, only Greenwood et al used availability of a support person after discharge.¹³ A couple of studies showed that proxies for socioeconomic status (not having a high school degree, being enrolled in Medicaid, and living in a poor neighborhood) were strong predictors of readmission.^{35 44}

Table 2 | Characteristics of patients and hospitals in studies that included specific patient populations

Study	Study population	Hospital type	Multicenter	Total sample size		Observed readmission rate (%)
				Derivation	Validation	
Asche et al, 2016 ²⁵	18+ AML patients	Academic and non-academic	Yes	3058	Fivefold cross validation	8.9
Benuzillo et al, 2018 ²⁶	18+ CABG patients	Academic and non-academic	Yes	1693	896	9.15
Cheung et al, 2018 ²⁷	18+ heart failure patients	Non-academic/ large community	No	4711	2019 (validation and testing)	13
Eby et al, 2015 ²⁸	18+ type 2 diabetes patients	Academic and non-academic	Yes	52 070	Bootstrap resampling with 500 iteration	10
Flythe et al, 2016 ²⁹	18+ hemodialysis patients	Academic	No	349	Bootstrap resampling with 1000 iteration	32.1
Golas et al, 2018 ³⁰	18+ heart failure patients	Academic and non-academic	Yes	11 510	Bootstrap 10-fold cross validation	23
Hatipoglu et al, 2018 ³¹	18+ pneumonia patients	Academic	No	1295	393	25
Horne et al, 2016 ³²	18+ heart failure patients	Academic and non-academic	Yes	Total: 6079	Total: 2663	14.1
				Female: 3013	Female: 1318	12.5
				Male: 3066	Male: 1334	16.5
				External total: 5162		15.6
					External female: 2537	14.6
					External male: 2625	18.6
Karunakaran et al, 2018 ³³	18+ type 2 diabetes patients	Academic	No	44 203	Split sample	20.4
Mahajan et al, 2018 ³⁴	18+ heart failure patients	VA health center	Yes	1210	Bootstrap 10-fold cross validation	21.7
Makam et al, 2017 ³⁵	18+ pneumonia patients	Academic and non-academic	Yes	1463	Fivefold cross validation	13.6
McGirt et al, 2015 ³⁶	18+ low back surgery patients	Academic	No	1803	361	5.9
Nguyen et al, 2018 ³⁷	18+ AML patients	Academic and non-academic	Yes	826	Fivefold cross validation	13
Padhukasahasram et al, 2015 ³⁸	18+ heart failure patients	Non-academic/ large community	No	789	10-fold cross validation	54.4
Reddy et al, 2018 ³⁹	18+ lupus patients	Academic and non-academic	Yes	9457	70:30 split	17.2
Rubin et al, 2016 ⁴⁰	18+ type 2 diabetes patients	Academic	No	26 522	17 681	20.4
Rubin et al, 2017 ⁴¹	18+ type 2 diabetes patients admitted to hospital for cardiovascular conditions	Academic	No	4950	3219	20
Rumshisky et al, 2016 ⁴²	18+ psychiatric patients	Academic	No	3281	1406	22
Shameer et al, 2016 ⁴³	18+ heart failure patients	Non-academic/ large community	No	748	320	16.6
Taber et al, 2015 ⁴⁴	18+ kidney transplant patients	Academic	No	1147	Bootstrap cross validation	11
Wang et al, 2016 ²⁴	18+ heart failure patients	Academic	No	4548	Bootstrap cross validation	33
Xiao et al, 2018 ⁴⁵	18+ heart failure patients	NA	NA	3000	67:16:16 split for training, validation, and testing	NA
Zheng et al, 2015 ⁴⁶	18+ heart failure patients	NA	NA	1641	Fivefold cross validation	19.3
Zolbanian et al, 2018 ⁴⁷	18+ heart failure patients	Academic and non-academic	Yes	Heart failure: 32 350; COPD: 31 070	70:30 10-fold cross validation	NA

AMI=acute myocardial infarction; CABG=coronary artery bypass graft; COPD=chronic obstructive pulmonary disease; NA=not available.

The top predictors among all models mostly included healthcare encounter history (previous emergency or inpatient visits within three to six months before index hospital admission)^{11 14 15} and a variety of clinical data indicating the severity of the patient's condition during the index admission (low level of albumin or using a variety of constructed severity scores).^{12 21 22} As stated above, a few studies also showed that disability/functional status measures and socioeconomic status were strong predictors of readmission.^{8 17 20 28 33 44}

Predictive models

Supplementary table E summarizes characteristics of predictive models used in the included studies. Timing of the prediction is of paramount importance for institutions to operationalize these risk assessment tools for readmission. Most studies (n=23) predicted

readmission right before or at discharge. Ten studies did not report the timing of their predictions; the rest reported it as within 24 hours after admission (n=3),^{10 23 32} before admission (n=1),²⁰ or after discharge (n=3).^{30 33 34} Most of the studies (n=24) examined more than one predictive model and chose the model with the highest C statistic and fewest predictors. Although all models are presented, for the ease of representation we chose the model with the highest C statistic for each study.

Out of 41 included studies, 26 used multivariable Cox or logistic regression models. Different feature selection techniques such as stepwise variable selection (forward, backward, or backward-forward methods),²² univariate binary regression, and LASSO (least absolute shrinkage and selection operator) were used.²³

Fifteen studies used various machine learning methods, including Bayesian conditional probability,^{10 43} random forest,^{14 15 38 46 47} neural network,^{15 19 27 39} deep learning,⁴¹ AdoBoost,²² gradient boosting,^{30 47} natural language processing,^{30 36 42} and others.^{24 45 46} The most popular machine learning methods used were random forest and neural networks. Shrinkage methods (for both traditional and machine learning models) such as LASSO or machine learning algorithms such as Ado boosting were used to limit the number of features.

On average, the C statistics for machine learning and traditional regression models were 0.74 (standard deviation 0.06; 95% confidence interval 0.71 to 0.77) and 0.71 (0.07; 0.68 to 0.73), respectively. Although the mean C statistic was higher for machine learning models, the difference was not statistically significant (difference 0.03, 95% confidence interval -0.0 to 0.07). Furthermore, we did not find a significant difference between the C statistics for all patient (0.76, 0.72 to 0.79) and patient specific (0.72, 0.70 to 0.75) models (difference 0.03, -0.01 to 0.07).

A few studies used other comprehensive methods of model evaluation, such as the integrated discrimination index and net reclassification index. For example, by calculating the clinical utility of a predictive model for a given threshold, Walsh et al also used their findings to develop a model of clinical usefulness to evaluate the potential cost of mis-calibration and to measure the value of interventions aimed at reducing readmission.²³ Of all studies, eight (20%) reported sensitivity and specificity of the developed models,^{15 20 24 25 30 39 42 46} and seven (17%) reported positive and negative predictive values.^{15 20 24 25 30 32 44} Finally, five (12%) reported being implemented in the EMR system.^{8 13 14 24 32}

Three studies used natural language processing to extract additional psychosocial information such as suicidality or excessive alcohol consumption that otherwise were not available via structured EMR data.^{30 36 42} For example, Rumshisky et al used 1000 informative words to extract additional data from related clinical notes, which improved the C statistic from 0.75 in the base model to 0.78.⁴² In addition to structured EMRs, Golas et al used natural language processing for two types of unstructured data—physicians' notes and discharge summaries—to analyze data related to patients' social history and treatment during admission (allergic reactions, history of illness, intolerances and sensitivities).³⁰

Quality assessment

We assessed the quality and risk of bias of studies by using six variables, including accounting for missing values, validation method, type of validation (internal versus external and prospective versus retrospective), calibration (yes/no), and scope of readmission assessment (only at studied hospitals or in a larger geographic area) (supplementary table F). We used a few techniques to deal with missing values in EMRs: removing data with missing values from the analytic sample,^{10 11} creating a separate category for them,⁸

imputing their values,¹⁴ and considering missing laboratory results to be normal.²¹

As we included only validated models, most were of high quality with a relatively low risk of bias. However, only a few expanded the assessment of their models beyond basic C statistics to evaluate the clinical usefulness of the models.^{24 46} Fifteen (37%) studies calibrated their models.^{8 12 18 19 21 23 26 31 34-37 41}

Calibration techniques such as Hosmer-Lemeshow, plot scaling, and prevalence adjustment were used to make the model probabilities more similar to the probabilities of the population studied. However, most of these studies failed to report the number of patients in each risk group, so we were unable to estimate the average predicted readmission rate and observed-to-expected ratios for each model. Furthermore, most models measured readmission only among included hospitals instead of using a broader (regional) scope for readmission. Finally, all validations were done internally. Thus, we could not assess the generalizability and practical utility of the developed readmission risk assessment tools.

Discussion

In this systematic review, we reviewed 41 studies of the development and validation of predictive models of 30 day hospital readmission using electronic medical records. These models were developed to identify patients at high risk of readmission for whom coordinated discharge care might reduce the chance of early readmission. On average, the predictive ability of risk readmission models based on EMR data compared with that of previously published models using all other available datasets (administrative or survey data) has improved, from 0.67 to 0.74.⁵

Comparisons with other studies

Over the past few years, despite increasing use of "big data" and rich clinical information available via EMRs,^{4 48} and application of sophisticated machine learning methods, predicting risk of early hospital readmission with reliable accuracy has remained elusive. Hospital readmission is a complex and multidimensional problem, demanding to be better understood. Although inclusion of essential clinical data available in EMRs (such as vital signs, laboratory results, or complexity of the surgical procedure) increased the predictive ability of the models, some important clinical data were still not readily available in EMRs. For example, composite measures of severity such as the Braden Score (risk indicator for developing a pressure ulcer),²² Comorbidity Point Score (risk indicator of multimorbid severity),¹² and Laboratory Acute Physiology Score (risk indicator of illness severity)^{12 21} have rarely been examined. Furthermore, functional status or frailty at the time of discharge, known to be an important risk factor for readmission,⁴⁹ is not routinely collected and used in EMR based predictive models.^{50 51}

Most notably, despite a large body of literature showing significant links between social and environmental factors and risk of readmission or other

adverse health events,^{52 53} health systems are still not systematically collecting these data. Including selected social and environmental factors,^{54 55} such as care giver availability or housing instability,⁵⁶⁻⁵⁹ could likely substantially improve the predictive accuracy of the risk readmission models.⁵⁵ To fill this void, alternative approaches have been examined. For instance, Census Bureau zip code or block level socioeconomic data have been merged with EMRs.^{30 33 35} Perhaps because of the imprecise nature of these aggregate data, however, many of them did not show significant difference in discriminatory power when examined in predictive models. Many models started using natural language processing to extract key social and environmental data from unstructured data such as physicians' notes.⁶⁰ Although physicians' or nurses' notes are unsystematically recorded, meaning that what is recorded by one physician may not be recorded by another, natural language processing has shown promising results for improving the accuracy of predictive models. Natural language processing can also be used to collect other salient information that is usually missing from structured EMRs, such as psychosocial or sensory statuses.^{30 36 42}

Additionally, the quality and integrity of EMR data are of concern and have implications for leveraging these data to develop accurate and precise risk assessment.⁶¹ This systematic review emphasized that nowhere in the literature was appropriate validation of EMR data performed. Furthermore, for missing data elements, standard approaches to identify the missingness mechanism and to appropriately deal with it without compromising the data elements used for further modeling are absent. Also, as previously described, EMRs lack certain salient data elements informed by the literature for risk assessment. We hypothesize that overcoming these data quality and integrity problems would not only improve the C statistics but would also introduce significant and impactful features that are missing from currently reviewed models.

Furthermore, with the emergence of big data and sophisticated machine learning methods in health-care, the number of predictive models of hospital readmission has increased over the past few years.⁶² The clinical utility of machine learning methods, however, needs further attention. For example, sophisticated machine learning methods such as neural networks work like a black box, lacking transparency in selection of features. Thus, their relative contribution, usefulness, and interpretability in medicine need to be investigated.⁶³ For example, class imbalance—when dichotomized outcomes are substantially different in probabilities—might cause biased predictions when machine learning is used, unless certain adjustments are made to correct it.⁶⁴ Otherwise, models that are developed using imbalanced training data would intrinsically provide more accurate predictions for the class with a higher number of occurrences. Despite extreme class imbalance in hospital readmission, only a small number of reviewed predictive models adjusted for it.

EMRs encompass a large repository of multi-dimensional data. Although traditional regression is easy to use and implement, it may not take advantage of the volume of data elements available in the EMR; however, machine learning methods are capable of using the exhaustive set of data elements for consideration.

Despite the growing literature supporting machine learning methods as an alternative, coupled with potential benefits in their use for predicting readmissions, three important remaining criteria have been highlighted, which our systematic review attempts to tackle. Firstly, feature selection remains an important criterion that is predicated on having an exhaustive and diverse set of data elements available, such as socioeconomic and functional status. Subsequent studies should consider implementing sufficiently granular data elements via text mining, merging this with smaller geographic units of analysis (census tract or neighborhood level), or encouraging health systems to collect these salient attributes. Secondly, machine learning methods struggle to achieve parsimony owing to the selection of several hundred to thousands of features to predict an outcome. The use of machine learning methods, although fashionable and offering a potential academic exercise, fails to answer important clinical questions about the implementation and interpretability of the results. Thirdly, machine learning methods vary substantially in their interpretability, creating barriers and impediments for clinical buy-in and for their implementation across health systems. Although interpretable machine learning methods have been absent in this systematic review, the evolution of the field requires development and implementation of interpretable machine learning methods to establish clinical usefulness and inspire potential changes in practice patterns.

The paucity of studies (5%) that provide information on the implementation and clinical utility of these models in the hospital setting leads to a substantial void in how these models can improve care coordination and discharge planning across readmission risk strata. With interpretable models that would enhance clinician “buy-in,” their implementation would encourage identification of patients who need efficient allocation of limited available resources for care coordination. Furthermore, these models would inform hospitals to tailor appropriate discharge protocols to the patients across readmission risk groups.

Regardless of the method used for prediction, careful diagnostic tests such as C statistics, sensitivity and specificity, positive and negative predictive values, integrated discrimination index, and net reclassification index should be calculated and discussed to ensure not only the accuracy of a model but also its clinical usefulness.⁶⁵ The C statistic is a measure of “discrimination” because it measures whether a model can discriminate patients at higher risk from those at lower risk. Besides C statistics, most of the reviewed models failed to calculate and interpret a reasonable array of other diagnostic tests

or even clinical usefulness of the models developed. Finally, to ensure approximate closeness of the model's performance to existing probabilities of the target population, a prediction model, regardless of how it was developed, needs to be well calibrated.²³ Most of the models we reviewed either did not discuss calibration or simply used goodness-of-fit tests such as Hosmer-Lemeshow in place of full calibration. Categorization into tenths of predicted counts and observed counts can be difficult if discrimination is poor.⁶⁶

Limitations of study

Our study had a few limitations. Most probably, the definition and classification of variables (both predictors and outcomes) varied among models. Although improving, EMR data are not yet standardized like administrative claims or survey data. As these studies used center specific EMRs, predictive models were developed for particular hospital settings and are not generalizable at a national level; therefore, the broader clinical and practice benefit may be localized and most likely applicable for institutional quality improvement. However, if disparate EMR data can be recoded in a manner that harmonizes across different EMRs, then comparisons can be made. Data coordinating centers could plausibly guide hospitals to do this, and this should be a future direction. We attempted to synthesize the findings to the best of our ability, so future studies may benefit from testing the variables and methods that were found to be most promising. Furthermore, few rigorous studies that have studied interventions targeting reduction in readmissions have actually shown a decrease in readmissions.^{67 68} Finally, as discussed above, most of the reviewed models neither included other recommended diagnostic tests besides C statistics nor discussed the clinical usefulness of their findings. To minimize bias, we chose only the highest quality models by including the ones that explicitly validated their findings.

Conclusions and policy implications

In short, despite notable progress in the development and accuracy of the models, predicting and reducing readmission remain a complex process.^{67 68} Most of the models developed to date have moderate predictive ability. No well accepted threshold of what constitutes an accurate C statistic exists because model discrimination is also a measure of how predictable a given outcome is. For an outcome such as 30 day readmission that is considered to be difficult to predict, a C statistic of 0.75 may be adequate for the model to be useful. In contrast, for an outcome that is readily predictable by clinical experts, even a model with a C statistic of 0.90 may not be useful. Use of EMR data and machine learning methods have created an enormous opportunity for further refinement of risk prediction tools for readmission, making them specifically pragmatic for hospitals to better identify patients who are at higher risk of readmission. Continued

development of these models to optimize performance of the model (tuning) may lead us toward improvement through institutional quality improvement and readmission reduction.

AUTHOR AFFILIATIONS

¹Department of Family Medicine, University of Michigan Medical School, Ann Arbor, MI, USA

²Institute for Healthcare Policy and Innovation, University of Michigan Medical School, Ann Arbor, MI, USA

³Department of Obstetrics and Gynecology, University of Michigan Medical School, Ann Arbor, MI, USA

⁴Department of Surgery, University of Michigan Medical School, Ann Arbor, MI, USA

⁵Department of Emergency Medicine, University of Michigan Medical School, Ann Arbor, MI, USA

⁶Undergraduate Research Opportunity Program, University of Michigan, Ann Arbor, MI, USA

⁷Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, USA

⁸Division of Gastroenterology and Hepatology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

⁹Michigan Integrated Center for Health Analytics and Medical Prediction (MiCHAMP), University of Michigan, Ann Arbor, MI, USA

¹⁰Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, USA

¹¹Veterans Affairs Center for Clinical Management Research, Ann Arbor, MI, USA

We acknowledge the tremendous help from Whitney Ann Townsend, Liaison Services Librarian to the Department of Family Medicine, in developing the initial search terms, and Murphy Vandervest and Shivani Shant from the Undergraduate Research Opportunity Program at the University of Michigan in screening of the initial search results. We thank Lois Phizacklea for her help in formatting the manuscript.

Funding: This study was supported by grants from the National Institutes of Health, P30 AG015281, and the Michigan Center for Urban African American Aging Research and from the University of Michigan Claude D Pepper Older Americans Independence Center, AG024824.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: support for the submitted work as described above; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work

Contributors: EM designed the data collection, developed the search strategy, monitored the data collection, screened and reviewed the selected articles, and drafted and revised the paper. NK drafted sections of the paper and revised the paper. NK developed and conducted the search strategy and drafted sections of the paper. GG screened the initial search results, reviewed selected studies, and drafted sections of the paper. AKW drafted and revised the paper. KS revised the paper. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. EM is the guarantor.

Ethical approval: Not applicable.

Data sharing: No additional data available.

Transparency: The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Dissemination to participants and related patient and public communities: Not applicable.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

1 Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. *JAMA* 2013;309:342-3. doi:10.1001/jama.2012.94856

- 2 Zuckerman RB, Sheingold SH, Orav EJ, Ruhter J, Epstein AM. Readmissions, Observation, and the Hospital Readmissions Reduction Program. *N Engl J Med* 2016;374:1543-51. doi:10.1056/NEJMsa1513024
- 3 Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019;110:12-22. doi:10.1016/j.jclinepi.2019.02.004
- 4 Jha AK, DesRoches CM, Campbell EG, et al. Use of electronic health records in U.S. hospitals. *N Engl J Med* 2009;360:1628-38. doi:10.1056/NEJMsa0900592
- 5 Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA* 2011;306:1688-98. doi:10.1001/jama.2011.1515
- 6 Zhou H, Della PR, Roberts P, Goh L, Dhaliwal SS. Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review. *BMJ Open* 2016;6:e011060. doi:10.1136/bmjopen-2016-011060
- 7 Debray TP, Damen JA, Snell KI, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;356:i6460. doi:10.1136/bmj.i6460
- 8 Amarasingham R, Velasco F, Xie B, et al. Electronic medical record-based multicondition models to predict the risk of 30 day readmission or death among adult medicine patients: validation and comparison to existing models. *BMC Med Inform Decis Mak* 2015;15:39. doi:10.1186/s12911-015-0162-6
- 9 Brindise LR, Steele RJ. Machine learning-based pre-discharge prediction of hospital readmission [abstract]. 2018 International Conference on Computer, Information and Telecommunication Systems (CITS). 2018. <https://www.researchgate.net/publication/325301578>.
- 10 Chen JH, Podchyska T, Altman RB. OrderRex: clinical order decision support and outcome predictions by data-mining electronic medical records. *J Am Med Inform Assoc* 2016;23:339-48. doi:10.1093/jamia/ocv091
- 11 Damery S, Combes G. Evaluating the predictive strength of the LACE index in identifying patients at high risk of hospital readmission following an inpatient episode: a retrospective cohort study. *BMJ Open* 2017;7:e016921. doi:10.1136/bmjopen-2017-016921
- 12 Escobar GJ, Ragins A, Scheirer P, Liu V, Robles J, Kipnis P. Nonelective rehospitalizations and postdischarge mortality: predictive models suitable for use in real time. *Med Care* 2015;53:916-23. doi:10.1097/MLR.0000000000000435
- 13 Greenwood KL, LaMori JC, Smith B, Doshi D, Davis C. Impact of behavioral health screening on proactive identification of patients at risk for hospital readmission. *Popul Health Manag* 2019;22:196-204. doi:10.1089/pop.2018.0074
- 14 Hao S, Wang Y, Jin B, et al. Development, validation and deployment of a real time 30 day hospital readmission risk assessment tool in the Maine healthcare information exchange. *PLoS One* 2015;10:e0140271. doi:10.1371/journal.pone.0140271
- 15 Jamei M, Nisnevich A, Wetchler E, Sudat S, Liu E. Predicting all-cause risk of 30-day hospital readmission using artificial neural networks. *PLoS One* 2017;12:e0181173. doi:10.1371/journal.pone.0181173
- 16 Logue E, Smucker W, Regan C. Admission data predict high hospital readmission risk. *J Am Board Fam Med* 2016;29:50-9. doi:10.3122/jabfm.2016.01.150127
- 17 Morris MS, Graham LA, Richman JS, et al. Postoperative 30-day Readmission: Time to Focus on What Happens Outside the Hospital. *Ann Surg* 2016;264:621-31. doi:10.1097/SLA.0000000000001855
- 18 Nguyen OK, Makam AN, Clark C, et al. Predicting all-cause readmissions using electronic health record data from the entire hospitalization: Model development and comparison. *J Hosp Med* 2016;11:473-80. doi:10.1002/jhm.2568
- 19 Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018;1:18. doi:10.1038/s41746-018-0029-1
- 20 Shadmi E, Flaks-Manov N, Hoshen M, Goldman O, Bitterman H, Balicer RD. Predicting 30-day readmissions with preadmission electronic health record data. *Med Care* 2015;53:283-9. doi:10.1097/MLR.0000000000000315
- 21 Tabak YP, Sun X, Nunez CM, Gupta V, Johannes RS. Predicting readmission at early hospitalization using electronic clinical data: An early readmission risk score. *Med Care* 2017;55:267-75. doi:10.1097/MLR.0000000000000654
- 22 Tong L, Erdmann C, Daldalian M, Li J, Esposito T. Comparison of predictive modeling approaches for 30-day all-cause non-elective readmission risk. *BMC Med Res Methodol* 2016;16:26. doi:10.1186/s12874-016-0128-0
- 23 Walsh CG, Sharman K, Hripcsak G. Beyond discrimination: A comparison of calibration methods and clinical usefulness of predictive models of readmission risk. *J Biomed Inform* 2017;76:9-18. doi:10.1016/j.jbi.2017.10.008
- 24 Wang H, Cui Z, Chen Y, Avidan M, Abdallah AB, Kronzer A. Predicting hospital readmission via cost-sensitive deep learning. *IEEE/ACM Trans Comput Biol Bioinform* 2018;15:1968-78. doi:10.1109/TCBB.2018.2827029
- 25 Asche CV, Ren J, Kirkness CS, Kim M, Dong Y, Hippler S. A prediction model to identify acute myocardial infarction (AMI) patients at risk for 30-day readmission [abstract]. Proceedings of the Summer Computer Simulation Conference. 2016. <https://www.researchgate.net/publication/316939926>.
- 26 Benuzillo J, Caine W, Evans RS, Roberts C, Lappe D, Doty J. Predicting readmission risk shortly after admission for CABG surgery. *J Card Surg* 2018;33:163-70. doi:10.1111/jocs.13565
- 27 Cheung BLP, Dahl D. Deep learning from electronic medical records using attention-based cross-modal convolutional neural networks [abstract]. 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) 2018.
- 28 Eby E, Hardwick C, Yu M, et al. Predictors of 30 day hospital readmission in patients with type 2 diabetes: a retrospective, case-control, database study. *Curr Med Res Opin* 2015;31:107-14. doi:10.1185/03007995.2014.981632
- 29 Flythe JE, Katsanos SL, Hu Y, Kshirsagar AV, Falk RJ, Moore CR. Predictors of 30-day hospital readmission among maintenance hemodialysis patients: A hospital's perspective. *Clin J Am Soc Nephrol* 2016;11:1005-14. doi:10.2215/CJN.11611115
- 30 Golas SB, Shibahara T, Agboola S, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Med Inform Decis Mak* 2018;18:44. doi:10.1186/s12911-018-0620-z
- 31 Hatipoğlu U, Wells BJ, Chagin K, Joshi D, Milinovich A, Rothberg MB. Predicting 30-day all-cause readmission risk for subjects admitted with pneumonia at the point of care. *Respir Care* 2018;63:43-9. doi:10.4187/respcare.05719
- 32 Horne BD, Budge D, Masica AL, et al. Early inpatient calculation of laboratory-based 30-day readmission risk scores empowers clinical risk modification during index hospitalization. *Am Heart J* 2017;185:101-9. doi:10.1016/j.ahj.2016.12.010
- 33 Karunakaran A, Zhao H, Rubin DJ. Predischarge and postdischarge risk factors for hospital readmission among patients with diabetes. *Med Care* 2018;56:634-42. doi:10.1097/MLR.0000000000000931
- 34 Mahajan SM, Burman P, Newton A, Heidenreich PA. A validated risk model for 30-day readmission for heart failure. *Stud Health Technol Inform* 2017;245:506-10.
- 35 Makam AN, Nguyen OK, Clark C, et al. Predicting 30-day pneumonia readmissions using electronic health record data. *J Hosp Med* 2017;12:209-16. doi:10.12788/jhm.2711
- 36 McGirt MJ, Sivaganesan A, Asher AL, Devin CJ. Prediction model for outcome after low-back surgery: individualized likelihood of complication, hospital readmission, return to work, and 12-month improvement in functional disability. *Neurosurg Focus* 2015;39:E13. doi:10.3171/2015.8.FOCUS15338
- 37 Nguyen OK, Makam AN, Clark C, Zhang S, Das SR, Halm EA. is Predicting 30-Day Hospital Readmissions in Acute Myocardial Infarction: The AMI "READMITS" (Renal Function, Elevated Brain Natriuretic Peptide, Age, Diabetes Mellitus, Nonmale Sex, Intervention with Timely Percutaneous Coronary Intervention, and Low Systolic Blood Pressure) Score. *J Am Heart Assoc* 2018;7:e008882. doi:10.1161/JAHA.118.008882
- 38 Padhukasahasram B, Reddy CK, Li Y, Lanfear DE. Joint impact of clinical and behavioral variables on the risk of unplanned readmission and death after a heart failure hospitalization. *PLoS One* 2015;10:e0129553. doi:10.1371/journal.pone.0129553
- 39 Reddy BK, Delen D. Predicting hospital readmission for lupus patients: An RNN-LSTM-based deep-learning methodology. *Comput Biol Med* 2018;101:199-209. doi:10.1016/j.combiomed.2018.08.029
- 40 Rubin DJ, Handorf EA, Golden SH, Nelson DB, McDonnell ME, Zhao H. Development and validation of a novel tool to predict hospital readmission risk among patients with diabetes. *Endocr Pract* 2016;22:1204-15. doi:10.4158/E161391.OR
- 41 Rubin DJ, Golden SH, McDonnell ME, Zhao H. Predicting readmission risk of patients with diabetes hospitalized for cardiovascular disease: a retrospective cohort study. *J Diabetes Complications* 2017;31:1332-9. doi:10.1016/j.jdiacomp.2017.04.021
- 42 Rumshisky A, Ghassemi M, Naumann T, et al. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Transl Psychiatry* 2016;6:e921. doi:10.1038/tp.2015.182
- 43 Shameer K, Johnson KW, Yahi A, et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: A case-study using Mount Sinai heart failure cohort. *Pac Symp Biocomput* 2017;22:276-87. doi:10.1142/9789813207813_0027

- 44 Taber DJ, Palanisamy AP, Srinivas TR, et al. Inclusion of dynamic clinical data improves the predictive performance of a 30-day readmission risk model in kidney transplantation. *Transplantation* 2015;99:324-30. doi:10.1097/TP.0000000000000565
- 45 Xiao C, Ma T, Dieng AB, Blei DM, Wang F. Readmission prediction via deep contextual embedding of clinical concepts. *PLoS One* 2018;13:e0195024. doi:10.1371/journal.pone.0195024
- 46 Zheng B, Zhang J, Yoon SW, Lam SS, Khasawneh M, Poranki S. Predictive modeling of hospital readmissions using metaheuristics and data mining. *Expert Syst Appl* 2015;42:7110-20. doi:10.1016/j.eswa.2015.04.066
- 47 Zolbanin HM, Delen D. Processing electronic medical records to improve predictive analytics outcomes for hospital readmissions. *Decis Support Syst* 2018;112:98-110. doi:10.1016/j.dss.2018.06.010
- 48 Jha AK, DesRoches CM, Kralovec PD, Joshi MS. A progress report on electronic health records in U.S. hospitals. *Health Aff (Millwood)* 2010;29:1951-7. doi:10.1377/hlthaff.2010.0502
- 49 Meddings J, Reichert H, Smith SN, et al. The impact of disability and social determinants of health on condition-specific readmissions beyond Medicare risk adjustments: A cohort study. *J Gen Intern Med* 2017;32:71-80. doi:10.1007/s11606-016-3869-x
- 50 Ng TP, Niti M, Tan WC, Cao Z, Ong KC, Eng P. Depressive symptoms and chronic obstructive pulmonary disease: effect on mortality, hospital readmission, symptom burden, functional status, and quality of life. *Arch Intern Med* 2007;167:60-7. doi:10.1001/archinte.167.1.60
- 51 Hoyer EH, Needham DM, Atanelov L, Knox B, Friedman M, Brotman DJ. Association of impaired functional status at hospital discharge and subsequent rehospitalization. *J Hosp Med* 2014;9:277-82. doi:10.1002/jhm.2152
- 52 Gottlieb LM, Tirozzi KJ, Manchanda R, Burns AR, Sandel MT. Moving electronic medical records upstream: incorporating social determinants of health. *Am J Prev Med* 2015;48:215-8. doi:10.1016/j.amepre.2014.07.009
- 53 Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* 2013;309:1351-2. doi:10.1001/jama.2013.393
- 54 Calvillo-King L, Arnold D, Eubank KJ, et al. Impact of social factors on risk of readmission or mortality in pneumonia and heart failure: systematic review. *J Gen Intern Med* 2013;28:269-82. doi:10.1007/s11606-012-2235-x
- 55 Kangovi S, Grande D. Hospital readmissions--not just a measure of quality. *JAMA* 2011;306:1796-7. doi:10.1001/jama.2011.1562
- 56 Annema C, Luttik ML, Jaarsma T. Reasons for readmission in heart failure: Perspectives of patients, caregivers, cardiologists, and heart failure nurses. *Heart Lung* 2009;38:427-34. doi:10.1016/j.hrting.2008.12.002
- 57 Martin RC, Brown R, Puffer L, et al. Readmission rates after abdominal surgery: the role of surgeon, primary caregiver, home health, and subacute rehab. *Ann Surg* 2011;254:591-7. doi:10.1097/SLA.0b013e3182300a38
- 58 Doran KM, Ragins KT, Iacomacci AL, Cunningham A, Jubanyik KJ, Jenq GY. The revolving hospital door: hospital readmissions among patients who are homeless. *Med Care* 2013;51:767-73. doi:10.1097/MLR.0b013e31829fafbb
- 59 Jiang HJ, Boutwell AE, Maxwell J, Bourgoin A, Regenstein M, Andres E. Understanding patient, provider, and system factors related to Medicaid readmissions. *Jt Comm J Qual Patient Saf* 2016;42:115-21. doi:10.1016/S1553-7250(16)42014-3
- 60 Navathe AS, Zhong F, Lei VJ, et al. Hospital readmission and social risk factors identified from physician notes. *Health Serv Res* 2018;53:1110-36. doi:10.1111/1475-6773.12670
- 61 Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. *N Engl J Med* 2010;363:501-4. doi:10.1056/NEJMp1006114
- 62 Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 2016;375:1216-9. doi:10.1056/NEJMp1606181
- 63 Watson DS, Krutzinna J, Bruce IN, et al. Clinical applications of machine learning algorithms: beyond the black box. *BMJ* 2019;364:l886. doi:10.1136/bmj.l886
- 64 Japkowicz N, Stephen S. The class imbalance problem: a systematic study. *Intelligent Data Analysis* 2002;6:429-49. doi:10.3233/IDA-2002-6504
- 65 Šimundić AM. Measures of diagnostic accuracy: Basic definitions. *EJIFCC* 2009;19:203-11.
- 66 Finazzi S, Poole D, Luciani D, Cogo PE, Bertolini G. Calibration belt for quality-of-care assessment based on dichotomous outcomes. *PLoS One* 2011;6:e16110. doi:10.1371/journal.pone.0016110
- 67 Dhalla IA, O'Brien T, Morra D, et al. Effect of a postdischarge virtual ward on readmission or death for high-risk patients: a randomized clinical trial. *JAMA* 2014;312:1305-12. doi:10.1001/jama.2014.11492
- 68 Goldman LE, Sarkar U, Kessel E, et al. Support from hospital to home for elders: a randomized trial. *Ann Intern Med* 2014;161:472-81. doi:10.7326/M14-0094

Web appendix: Supplementary tables