

A Systematic Review of Highly Generalizable Risk Factors for Unplanned 30-Day All-Cause Hospital Readmissions

Peng Zhao¹ and Illhoi Yoo^{1,2*}

¹Informatics Institute, University of Missouri, Columbia, MO, United States

²Department of Health Management and Informatics, School of Medicine, University of Missouri, Columbia, MO, United States

Abstract

Background: Hospital readmissions are common and expensive. Numerous global efforts have been devoted to predicting readmissions. However, for many reasons including the variations in the studied populations and inconsistent definitions of readmissions, the outcomes of some studies can hardly be generalized to other studies inside or outside the same country or region.

Objective: The objective was to identify highly generalizable risk factors for unplanned 30-day all-cause hospital readmissions to guide the selection of baseline predictor variables in different readmission studies.

Methods: In July 2017, PubMed was searched to identify articles pertaining to the risk factors for unplanned 30-day all-cause hospital readmissions. To identify potentially eligible risk factors, characteristics of the selected studies were manually extracted. The generalizability of the risk factors was assessed with predefined criteria.

Results: 13 articles were eligible for the review. A total of 42 risk factors were identified and 34 of them were found to be highly generalizable.

Conclusions: The 34 risk factors are not specific to any populations or places, and the corresponding predictor variables can serve as baseline variables in readmission prediction studies. No major difference has been observed between the risk factors identified inside and outside the United States except that US studies appeared to prefer composite comorbidity measures. All the reviewed studies have used traditional statistical regression-based methods to identify risk factors and more applications of data mining techniques are expected in this field.

Keywords: Unplanned 30-day hospital readmission; Patient readmission; Generalizable risk factors; All-cause

Introduction

Hospital readmissions have emerged as a global concern due to their high frequency and high associated cost. In the United States, 15.5% of Medicare beneficiaries were readmitted unplanned within 30 days after discharge during 2014/7 to 2015/6 [1]. It has been estimated that unplanned readmissions account for \$17.4 billion in Medicare expenditure annually [2]. In England, the 28-day emergency readmission rate was 11.5% during the fiscal year 2011 to 2012 [3]. Although the causes have not been well studied, unplanned hospital readmissions are frequently seen as related to the substandard quality of care received during index admissions. The pressure to reduce cost and improve healthcare quality has triggered the implementations of financial penalty programs targeting readmissions in different countries, including England, Germany, and the United States [4]. In the fiscal year 2017, 79% of the hospitals will be penalized by the Hospital Readmissions Reduction Program (HRRP) [5] in the United States with the total estimated penalties of \$528 million [6].

Recent years have seen a growing body of literature on hospital readmissions with the goal of improving healthcare quality and lowering cost. The predictive modelling of readmissions is one of the most common study types to help providers better identify high-risk patients. Unfortunately, studies in this area are highly fragmented, especially in target populations. The study outcomes span from models that are specific to populations with particular diseases or surgeries to general purpose all-cause models. Currently, the HRRP in the United States only considers the index conditions or surgeries of acute myocardial infarction, heart failure, pneumonia, chronic obstructive pulmonary disease, elective total hip or knee arthroplasty, and coronary artery bypass graft in the calculation of readmission penalties due to their high

prevalence and cost [5,7]. Largely spurred by the HRRP, many studies have focused on readmissions occurring after the index admissions for these conditions or surgeries only. The choice between condition-specific and all-cause readmission models has long been under debate. However, condition-specific models have been criticized for the poor generalizability, especially in patients with multiple conditions [8,9]. In addition, readmissions are not always clinically relevant to the index admissions. One retrospective analysis of 217,767 unplanned readmissions has found that 58.5% of them were not assigned the same principal diagnoses, diagnosis related groups (DRG), or all-patient DRG as the index admissions [10].

Attempts to predict readmissions were further complicated by the lack of consensus on data inclusion criteria. Many studies focused on unplanned readmissions while some others included all available readmissions without removing scheduled readmissions. The definitions of unplanned readmissions were also highly inconsistent. Some studies restricted unplanned readmissions to occur in certain departments or specialties and some identified them by diagnosis codes. In some

***Corresponding author:** Illhoi Yoo, PhD, Department of Health Management and Informatics, School of Medicine, University of Missouri, CE718 Clinical Support and Education Building (DC006.00), Columbia, MO, 65212, United States, Tel: 1 573 882 7642; Fax: 1 573 882 6158; E-mail: yooil@health.missouri.edu

Received August 30, 2017; **Accepted** September 11, 2017; **Published** September 15, 2017

Citation: Zhao P, Yoo I (2017) A Systematic Review of Highly Generalizable Risk Factors for Unplanned 30-Day All-Cause Hospital Readmissions. J Health Med Informat 8: 283. doi: [10.4172/2157-7420.1000283](https://doi.org/10.4172/2157-7420.1000283)

Copyright: © 2017 Zhao P, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

studies, unplanned readmissions were further classified as either potentially avoidable or unavoidable. Readmissions due to progressions of existing conditions or newly developed conditions after discharge are deemed unavoidable [11,12]. It has been argued that including unavoidable readmissions in quality measures is unfair because they are not directly related to the quality of healthcare services during index admissions [11]. However, there is no agreement at present on the criteria to identify avoidable readmissions. In many studies, the avoidable readmissions were determined by medical experts and the inclusion eligibility can be subjective [13]. According to a systematic review of 34 articles in 2011, the measured proportions of avoidable readmissions varied from 5% to 79% [11].

To exacerbate the situation, diverse time frames were used to capture readmissions. In a systematic review of 26 readmission prediction models developed in six countries, the intervals between discharges and readmissions ranged from 14 days to four years [14]. The Centers for Medicare and Medicaid Services (CMS) in the United States adopted the 30-day time window [5]. In the United Kingdom, both 28-day [15] and 30-day [16] periods were used by the National Health Service (NHS) to measure readmission rates. 30-day is currently the most used time frame globally. The possible reasons are that older patients are more vulnerable during this period [17] and readmissions occurring within 30 days are more likely influenced by the quality of care [18].

Given the complex nature of hospital readmissions, it is challenging to conduct meaningful readmission prediction studies without good knowledge of existing evidence from both domestic and global research communities. However, due to the heterogeneous target populations and inconsistent definitions of readmissions, the outcomes of some studies can be hardly generalized to other studies [19]. The purpose of this study was to identify the generalizable study outcomes of readmission predictions from the risk factor level to guide the selection of baseline predictor variables in different readmission studies. Especially, we are interested in the risk factors for unplanned 30-day all-cause hospital readmissions due to the better-validated time frame and the broader target.

In the past few years, several attempts have been made to review risk factors or predictor variables for hospital readmissions, yet none of them have focused on generalizability. The review by Vest et al. [20] in 2010 was limited to US studies from 2000 to 2009 only and the time frame varied from seven-day to six-month for all-cause readmissions. The focus of the review by Kansagara et al. [14] in 2011 was on readmission prediction models derived in developed countries before 2011. Predictor variables of the reviewed models were tabulated without differentiating condition-specific and all-cause models. Zhou et al. [21]

also placed emphasis on reviewing readmission prediction models developed between 2011 and 2015. The significant predictor variables were summarized without further analysis. Besides, many studies have reviewed readmission risk factors for specific conditions or surgeries. To the best of our knowledge, this is the first systematic review of highly generalizable risk factors for unplanned 30-day all-cause hospital readmissions.

Materials and Methods

Data source and search strategy

In this study, we present a systematic literature review of highly generalizable risk factors for unplanned 30-day all-cause hospital readmissions following the PRISMA statement for systematic reviews [22].

A literature search was performed in PubMed to identify articles relating to the risk factors for unplanned 30-day all-cause hospital readmissions. The search keywords have four components reflecting the interest of this review: “unplanned”, “30-day”, “hospital readmission”, “risk factors”. “All-cause” was not included in the keyword because some all-cause readmission studies do not explicitly mention their scopes. Synonyms and hyphenations were included to account for variations in different studies. Wildcards were used to match the verb and noun forms of “readmission”. The logical relationships among the search keywords are shown in Figure 1.

Study inclusion and exclusion criteria

In this study, only peer-reviewed articles written in English were considered. We included articles focusing on identifying statistically significant predictor variables or risk factors for 30-day unplanned all-cause hospital readmissions. Articles were excluded if they met any of the following criteria:

- (1) The readmission time frame is other than 30 days, such as 90-day readmission,
- (2) Studies focusing on planned readmissions or not differentiating planned and unplanned readmissions,
- (3) Studies specific to narrow patient populations with particular medical conditions or underwent certain surgeries,
- (4) The study outcome is more than 30-day unplanned all-cause hospital readmission, such as mortality in combination with 30-day unplanned all-cause hospital readmission,
- (5) Studies of paediatric and new-born readmissions. Paediatric and new-born readmissions were filtered out because the risk factors may be distinct from adult readmissions [14,20] and the readmissions could be influenced by parent-side factors [20,23].

```
(unplanned OR unexpected OR avoidable OR preventable)
AND
(thirty-day OR thirty day OR 30-day OR 30 day)
AND
(readmi* OR re-admi* OR patient readmission [MeSH Terms] OR rehospit* OR re-hospit*)
AND
(risk factors OR predictors OR determinants OR characteristics)
```

Figure 1: Logical relationships among the search keywords.

To reduce redundancy and bias, external validations of existing prediction tools were removed and only the original articles of the cited tools were included if eligible.

Data extraction process

The characteristics of the studies, including the publication year, study region, data source, study design, cohort definition and definition of unplanned readmissions, analysis method, predictor variables, and statistically identified risk factors ($P < 0.05$) were extracted from all the included studies. The risk factors were summarized by category and were grouped if they share the same corresponding predictor variable. The number of studies that analysed predictor variables and the number of studies that found them significant were recorded.

Generalizability assessment

In this study, we define “high generalizability” as the capability of being applied to other hospital readmission prediction studies regardless of target populations and residing places. Although the studies specific to narrow populations have been filtered out during the article selection step, some identified risk factors may be still tied to a sub-population. For example, studies with Medicare patients are less generalizable because those patients are 65 years old or older in the United States. Also, some risk factors identified in one place may not work in other places if they are closely related to the unique local health care systems (e.g., insurance, medical social welfare). In addition, it may be impractical to apply some risk factors to other types of studies due to the difference in study exposures related to designs. As a result, we chose to assess the generalizability of risk factors by three questions:

- (1) Whether a risk factor is specific to a narrow population or not,
- (2) Whether it is specific to a place or not,
- (3) Whether it is specific to a study exposure related to one particular study design or not.

If a risk factor is not specific to any of them, we deem that the risk factor is generalizable.

Results

Study selection

Figure 2 shows the process of identifying eligible articles. The initial query was performed on July 21, 2017, and returned 370 articles. After removal of one duplicated article and two non-English articles, the remaining 367 articles were reviewed based on titles and abstracts. 331 of them met the exclusion criteria and were filtered out. The remaining 36 articles were then reviewed in full text. One article was removed because it is not a peer-reviewed article. Four studies were excluded from the list because they are external validations of two existing readmission prediction models without major modifications (two articles validated the HOSPITAL score [24], one article validated the LACE score [25], and one article validated both the HOSPITAL and LACE scores). The original article of HOSPITAL score was included in this review while the LACE score article was not because the study outcome was both mortality and 30-day unplanned readmission. Four articles were specific to certain diagnoses and thus were removed. Five studies were filtered out because they did not differentiate unplanned readmissions. Nine articles were excluded because they did not report statistically significant predictor variables or risk factors. The remaining 13 highly relevant articles were included in this literature review.

Data extraction

Table 1 in the appendix summarizes the characteristics of the 13 studies. The literature on this topic is very recent. Although we did not intentionally limit the publication date, the earliest eligible article was published in 2009, reflecting a growing interest in predicting 30-day unplanned all-cause readmissions in the past decade. Of the 13 studies, over half (7/13) are based in the United States, two in Israel, two in Singapore, one in Sweden, and one in Taiwan. The majority (12/13) of the studies are retrospective and only one study [26] adopted the prospective design. Multivariate logistic regression is the most used analysis method (12/13) to identify significant predictor variables and only one study [27] used Poisson regression.

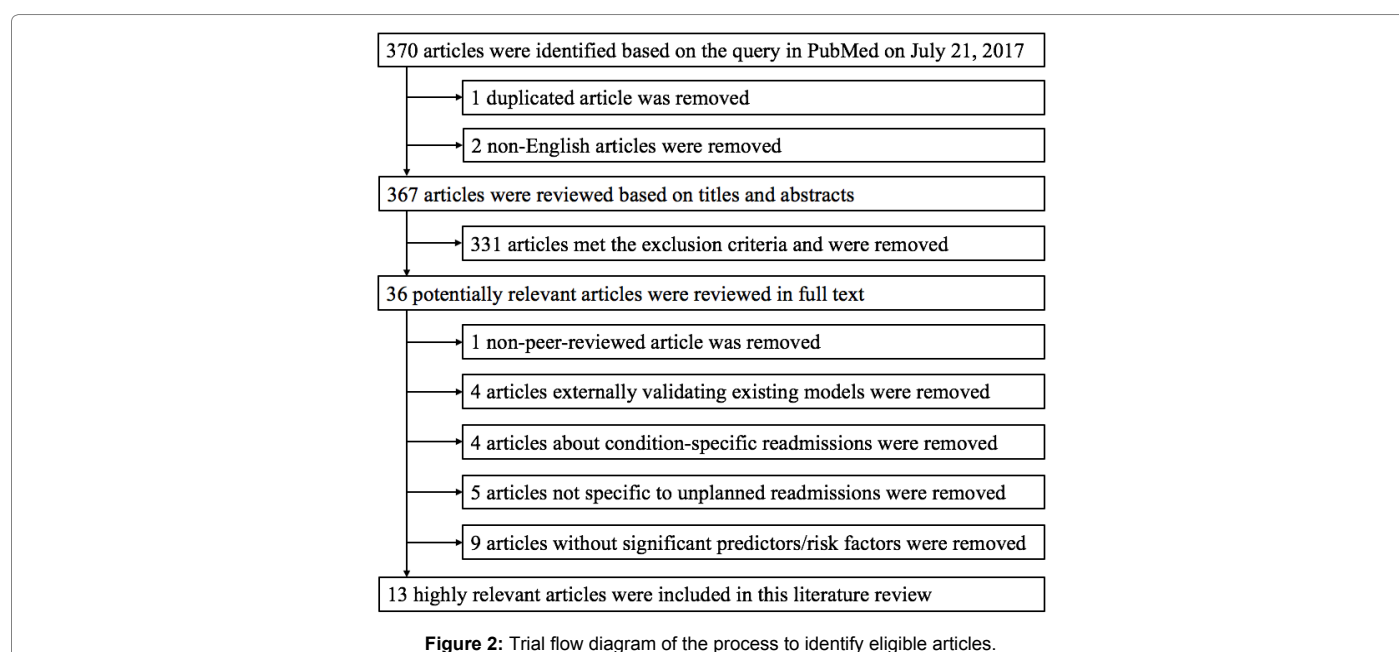


Figure 2: Trial flow diagram of the process to identify eligible articles.

Categories	Predictor variables	# Significant / # analyzed		
		Total	US	Non-US
Sociodemographic factors	Age [^] ~	5/11	2/6	3/5
	Gender [^] ~	2/10	1/6	1/4
	Race or ethnicity ~	2/6	2/4	0/2
	Rurality ~	2/4	2/3	0/1
	Insurance payer*	2/2	2/2	
	Education ~	1/3	0/1	1/2
	Admission class*	1/1	-	1/1
	Required financial assistance*	1/1	-	1/1
	Homelessness	1/1	1/1	
	Index admission in a subsidized ward*	1/1		1/1
Healthcare utilizations	Number of hospital admissions [^] ~	7/7	3/3	4/4
	Number of emergency department visits [^] ~	2/2		2/2
	Home care services	1/1		1/1
	Nursing home resident	1/1		1/1
Index admission characteristics	Length of stay [^] ~	5/8	4/5	1/3
	Admission type [^] ~	2/2	1/1	1/1
	Admission itself is a readmission	1/1	1/1	
	Discharged from oncology service	1/1	1/1	
	Required inpatient dialysis	1/1		1/1
	Required procedures	1/1	1/1	
Comorbidities and conditions	Comorbidity indices [^] ~	3/5	2/4	1/1
	Number of comorbidities	2/3	2/3	
	Cancer/malignancy [^] ~	2/3		2/3
	Anemia ~	1/2		1/2
	Chronic obstructive pulmonary disease ~	1/2		1/2
	Depression ~	1/2	0/1	1/1
	Diabetes mellitus ~	1/2		1/2
	Heart diseases ~	1/2		1/2
	Acute kidney injury	1/1	1/1	
	Chronic renal failure	1/1		1/1
	Chronic kidney disease	1/1		1/1
	Malnutrition	1/1		1/1
	Sepsis	1/1	1/1	
Lab tests	Hemoglobin level at discharge	2/2	2/2	
	Albumin level ~	1/2	0/1	1/1
	Sodium level at discharge	1/1	1/1	
Medication	Treatment with anti-depressants	1/1		1/1
Functional status and health literacy	At-admission activities of daily living*	1/1		1/1
	In-hospital activities of daily living decline*	1/1		1/1
	Health literacy*	1/1	1/1	
Hospital factors	Bed occupancy	1/1		1/1
	Admitted to a Veterans Affairs hospital*	1/1	1/1	

[^]Predictor variables found to be significant in more than one countries or regions
[~]Predictor variables studied in more than one countries or regions
^{*}Predictor variables with low generalizability

Table 1: Summary of the corresponding predictor variables of the identified risk factors.

The studies are highly heterogeneous in data type and data sources. Two studies [28,29] used claims data only and five studies [24,30-33] used clinical and/or administrative data from electronic health records (EHR) only. Four studies [26,34-36] combined data from various sources, including proprietary EHR, validated questionnaires, hospital information systems, veterans affairs database, and Medicare dataset. One article [37] studied state-level discharge summary data and one study [27] retrospectively analysed the control group of a clinical trial.

The definitions of unplanned readmissions are also very distinct. Four studies [26,32,33,37] directly used the data of unplanned readmissions without any definitions. Two studies [31,35] only included readmissions to emergency departments within 30 days of discharge because emergency department visits are not scheduled in advance. Three studies [28,30,36] excluded planned readmissions based on Clinical Classification Software (CCS) [38] or Diagnosis Related Group (DRG) [39] codes, including transplantations, psychiatric issues, maintenance chemotherapy, dental procedures, pregnancy-related procedures, and other planned procedures. Two

studies [27,34] excluded admissions to the specialties of obstetrics, gynaecology, dentistry, otolaryngology, ophthalmology, orthopaedic surgery, general surgery, or psychiatry. One study [29] excluded admissions with a principal diagnosis of cancer because cancer patients may have planned stays for cancer treatments. One study [24] separated readmissions into potentially avoidable and unavoidable based on administrative data with a validated algorithm SQLape [40]. Unavoidable readmissions include planned readmissions and any unforeseen readmissions for new conditions not related to known diseases during the index admissions [24]. The unavoidable readmissions were excluded from the analysis.

From the 13 studies, a total of 42 risk factors were identified and their corresponding predictor variables were aggregated and summarized in Table 1. They belong to eight major categories, including sociodemographic factors, healthcare utilizations, index admission characteristics, comorbidities and conditions, lab tests, medication, functional status and health literacy, and hospital factors. For each predictor variable, the number of studies found it significant was reported along with the number of studies included it in the analysis. 13 predictor variables were found to be statistically significant ($P < 0.05$) in more than one studies (including age, gender, race, rurality, the insurance payer, the number of hospital admissions in six months or one year before the index admission, the number of emergency department visits in six months or one year before the index admission, the length of stay of the index admission, the type of the index admission, the comorbidity indices, the number of comorbidities, cancer, and the haemoglobin level at discharge). 17 predictor variables were studied in more than one countries or regions and eight of them were found to be significant ($P < 0.05$) in more than one countries or regions (including age, gender, the number of hospital admissions in six months or one year before the index admission, the number of emergency department visits in six months or one year before the index admission, the length of stay of the index admission, the type of the index admission, the comorbidity indices, and cancer). Of the 42 risk factors, 34 meet our generalizability requirements (with answers NO to the three questions) and were found to be highly generalizable. The corresponding predictor variables of the eight risk factors with low generalizability were labelled with asterisk (*) in Table 1 (including the insurance payer, required financial assistance, index admission class, index admission was in a Veterans Affairs hospital, index admission was in a subsidized ward, at-admission activities of daily living, in-hospital activities of daily living decline, and health literacy).

Although it was not the intention of this study to review risk factors only applicable to the United States, about half (7/13) of the studies were based in the United States. To account for the potential bias towards US studies, it is meaningful to compare the risk factors identified within and outside the United States. For each variable, the number of studies found it significant and the number of studies analysed it were further classified by study regions (either US or non-US) (Table 1). For predictor variables only studied in one region, the corresponding numbers in another region were left blank for the sake of clarity. No obvious regional difference was observed for the eight categories, except that the studies in the United States preferred composite comorbidity measures (comorbidity indices and the number of comorbidities) to the presence of individual comorbidities. However, this cannot be justified by significance tests due to the small sample size.

Discussion

From the 13 studies, 42 risk factors have been identified with 34 being highly generalizable. Their rationale, generalizability, and identification methods will be discussed in this section.

Sociodemographic factors

In this review, sociodemographic factors were reported by most studies. Age, gender, race, and socioeconomic status are normally used as predictor variables to account for demographic and social influences on readmissions.

Older age has been reported to associate with higher readmission rates [19,41]. The possible reason is that older patients are often frailer and face more health issues than younger patients, such as comorbidities and polypharmacy [19]. Studies have also observed significant differences in readmission rates between genders [42-44]. Besides biological differences, gender-related social behaviours may play a role in the different readmission patterns [45]. Race and ethnicity can also potentially affect readmissions because they are dimensions of a society's stratification system to distribute resources, risks, and rewards [45].

Socioeconomic status measures an individual or a group's economic and social position by considering income, education, and occupation [46]. Evidence showed that poor physical and psychological health outcomes, including hospital readmissions, were associated with socioeconomic status disadvantage (e.g., low income, limited education, substandard neighbourhood) [47-49]. Although the mechanism is still under debate, lower socioeconomic status was reported to indirectly affect health by causing more stress, exposure to worse physical or social environments, unhealthy lifestyles, or limited access to healthcare resources [50].

Age was considered in 11 studies among which five studies found that increasing age or older age were significantly associated with readmissions. Two studies found that male gender was a risk factor. African American race was found to have a higher readmission risk in two studies. Living in a rural area, having certain insurance payers, education level lower or equal to high school, requiring medical financial assistance are other reported risk factors.

It is worth noting that some factors under this category may depend on or interact with each other. One example is that, in the United States, most people need to reach age 65 to qualify for Medicare, a national insurance program administered by the US government [51]. In this case, Medicare insurance depends on age. These factors can further interact with each other in more implicit ways. Therefore, studies with these factors may need more careful planning and design.

In addition, factors in this category are unmodifiable. It has long been argued in the United States that using readmission rate as a quality indicator without adjusting for unmodifiable socioeconomic factors is unfair because they are beyond the control of hospitals [52]. In 2016, the socioeconomic risk adjustment in hospital readmission measures was finally enforced by the "21st Century Cures Act" [53].

Healthcare utilizations

Many studies have incorporated patients' previous healthcare utilizations into readmission prediction models. The assumption is that higher utilizations such as repeated admissions to hospitals or emergency department's visits prior to the index admissions may account for the total burden of illness [24], which can potentially relate to readmissions.

Six months or one year are the most common look back periods to count previous hospital admissions or emergency department visits. A longer look-back period may potentially include utilizations less relevant to the readmission of interest and dilute the impact of more recent utilizations. Besides higher numbers of previous hospital admissions and emergency department visits, “received home care services” and “being a nursing home resident” were also identified to associate with higher readmission risks.

Index admission characteristics

It has been shown that the length of stay of index admissions can influence readmissions [54]. A longer stay may indicate a more complicated underlying situation and may expose the patient to more risks [13]. However, a shorter stay may also link to a higher readmission risk because the patient may not be ready for early discharge [13,55]. The relationship between the readmission risk and the length of stay has been found to be U-shaped rather than monotonic [54]. In this review, we did not observe a large discrepancy in the effect of length of stay between the studies as they all agreed that longer index admissions were related to higher risk of readmissions.

Besides the length of stay, the risk factors of acute admission type, admission is a readmission, discharged from oncology service, required inpatient dialysis, and required procedures during the index admission all indicate that patients were in severe situations during the index admissions.

Comorbidities, conditions, lab tests, and medications

It is well established that comorbidities are associated with undesired healthcare outcomes [56-58]. To date, there has been no consensus on the definition of comorbidity yet, but the core concept is the coexistence of more than one condition in the same patient [59]. Evidence shows that the top primary diagnoses of potentially avoidable readmissions are often possible complications of a comorbidity [60] and higher comorbidity has been linked to increased readmission risks [61,62]. In readmission predictions, comorbidities are represented either in the form of the number of comorbidities, the comorbidity index, or the presence of a comorbid condition.

It has been found that the readmission risk will rise as the number of comorbidities increases from the reviewed studies. More than just counting the number of comorbidities, the comorbidity index further accounts for contributions of different comorbidities. Charlson [63] and Elixhauser [64] are the most commonly used comorbidity indices [65]. The Charlson index was originally developed based on medical record review of 19 comorbid conditions with each condition assigned a weight of 1, 2, 3, or 6 depending on the risk associated with mortality [63]. A higher total index indicates a greater chance of one-year mortality. The Charlson/Deyo index is a highly referred variant by adapting the original index to 17 categories of comorbid conditions with ICD-9-CM codes [66]. The Elixhauser index includes a more comprehensive list of 30 comorbidities [64] but with little overlap with the Charlson index [65]. According to a systematic review of 54 articles in 2012, the Elixhauser index generally outperforms other available indices [65].

The presences of some chronic or acute conditions are also related to readmissions. Especially, cancer, chronic obstructive pulmonary disease, heart diseases, renal diseases, diabetes mellitus, and sepsis found in this review are among conditions associated with the most readmissions [7]. The included lab tests and medication are closely related to some conditions on the list, such as Anaemia, renal diseases, and depression.

It has been argued that inclusion of comorbidity measures or diagnosis codes in readmission prediction models may reduce the timeliness of the predictions. The reason is that in practice they are normally only available after discharge [24].

Functional status and health literacy

According to Leidy’s definition, functional status measures a person’s ability to provide for the necessities of life, including daily activities to meet basic needs, fulfil usual roles, and maintain health and well-being [67]. The impairment of functional status has been reported to associate with increased risk of readmissions [68].

Health literacy was defined as “the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions” by Ratzan et al. in 2000 [69] and this definition was adopted by the Institute of Medicine (IOM) of the United States [70]. Although not considered as a social factor, it is more distally influenced by social factors [45]. Low health literacy may attribute to no adherence to treatment plans, compromised communications with clinicians, limited self-care skills [71], and is associated with many poor health outcomes, including hospital readmissions [72]. To assess health literacy, questionnaire-based tests are administered and several tools are available [73].

Evidence showed the inclusion of functional status or health literacy can increase the predictive performance of readmission models [26]. However, they are seldom used due to the difficulty of data collection [24], especially in the case of retrospective studies.

Hospital factors

The factors from the hospital side may also contribute to readmissions in many ways. For example, the pressures in hospital resources (e.g., beds) may cause premature discharges of existing patients, which have shown to be related to readmissions [55]. There is also evidence that medical errors associate with higher readmission risks [74].

However, similar to the finding of another study [20], most of the identified risk factors are patient-side factors or clinical factors and only two hospital-side risk factors (inpatient bed occupancy>95%, index admission was in a Veterans Affairs Hospital) were found. This could be attributed to the small sample size, but a more plausible reason is that most studies followed the single-center retrospective cohort design. For single-center retrospective studies, it is harder to collect encounter-level hospital-side factors. If possible, it is recommended to collect multi-center data or combine with other data sources, such as claims data, to account for the variances in hospital-side factors.

Another possible reason is that, unlike patient-side factors and clinical factors, which are usually well-defined and readily available in administrative and clinical databases, hospital-side factors are harder to collect. More efforts are needed to define and quantify hospital-side factors in higher granularity beyond the basic hospital characteristics, such as geolocation, hospital type, teaching status, and beds, especially for studies measuring readmission rates for quality compare purposes.

Generalizability of the risk factors

The objective of this study was to review risk factors that can be widely generalized regardless of target populations and their residing countries. The generalizability of the 42 identified risk factors was assessed by the three questions detailed in the methods.

34 of the risk factors meet our generalizability criteria (with answers of NO to the three questions). The corresponding predictor variables of the eight risk factors with low generalizability include the insurance payer, required financial assistance, index admission class, index admission was in a Veterans Affairs hospital, index admission was in a subsidized ward, at-admission activities of daily living, in-hospital activities of daily living decline, and health literacy.

Health insurance is country-specific. “Medicare/Medicaid as insurance” and “Medi-Cal as insurance” are significant but they are only applicable in the United States. The insurance payer is a useless predictor variable for countries with universal health care coverage. “Requiring financial assistance from Medifund”, “index admission was in a subsidized ward”, and “index admission class A” may indicate a lower socioeconomic status but they all closely relate to the financial regulations and social welfare of the patients’ residing countries. The difficulty of collecting these data can be distinct in different countries. “Index admission was in a Veterans Affairs hospital” is only valid in the United States. We excluded functional status and health literacy because they are often harder to collect (e.g., interviews, self-reporting) for retrospective studies.

We kept comorbidity indices in the list of highly generalizable risk factors. Although the Charlson/Deyo and Elixhauser indices were originally built based on ICD-9-CM codes, which were the adaption of ICD-9 codes in the United States [75], they have been successfully translated to work with ICD-10 codes in Canada and Switzerland [76,77].

Methods to identify risk factors

The reviewed studies are highly consistent in analytical methods. 12 studies used logistic regression and one used Poisson regression. Logistic regression and Poisson regression both belong to the family of generalized linear models, which estimate model parameters by maximizing likelihood [78]. Poisson regression assumes the response variable follows a Poisson distribution, while in logistic regression the response variable can be binomial, ordinal, or multinomial. Binomial logistic regression is usually used in readmission predictions because the outcome is dichotomous (either readmitted or not readmitted). In binomial logistic regression, the binary response variable is linked to the linear combination of independent predictor variables through a logit function [79]. Poisson regression models a discrete count response variable with the logarithm as the link function [78]. In these studies, the adjusted odds ratio was the most used metric to assess a variable’s degree of association to the response variable. The odds ratio measures the relative chance of an outcome of interest to occur under different exposures [80]. The significance levels were set to 0.05 in all the studies.

Surprisingly, none of the 13 reviewed studies has used methods other than traditional statistical analysis. In recent years, data mining has been a hot research area and there have been many successful applications in healthcare [81]. Unlike statistics, which are hypothetic-deductive, data mining uses more flexible and more inductive ways to find patterns hidden in data [81]. Decision trees [82] are a family of supervised classifiers especially suitable to identify risk factors. The process to assign a label to the response variable can be visualized in a straightforward tree-like structure. The critical cut-off values of predictor variables associated with readmissions can be directly obtained from decision trees. The association rule mining [83] is another data mining technique appropriate to identify risk factors. This technique intends to discover strong rules (frequent item sets) from data based on predefined criteria. The risk factors can be extracted from the rules with high ranks.

Another concern is that some studies reported results without evaluations and/or internal validations of the prediction models. To reduce the bias and improve the usefulness of a prediction model, it is recommended to report prediction models following the guidelines in the TRIPOD statement [84] and the statement from the American Heart Association [85]. Especially, it is important to evaluate and report the model’s performance in the derivation and validation datasets.

The most popular model evaluation metric is the area under the receiver operating characteristic curve (AUC, or AUROC), or called the *c*-statistic [86]. The receiver operating characteristic curve is a graphical representation of a binary classifier’s performance as the discrimination threshold is varied [87]. The AUC measures the model’s ability of discrimination and can be interpreted as the probability that the model will rank a randomly selected positive sample higher than a randomly selected negative sample [87]. The AUC ranges from 0.5 to 1 with 1 indicating a perfect classifier.

The two widely used validation methods are hold-out cross-validation and k-fold cross-validation [88]. The hold-out method splits the dataset into a derivation dataset and a validation dataset. The derivation dataset is used to build the prediction model and the validation dataset is used to test the model. The disadvantage of the hold-out method is the partition of the original dataset might be biased and the resulting derivation and validation datasets might follow different local distributions. To overcome this issue, k-fold cross-validation method randomly splits the original dataset into *k* equal-sized partitions and uses one partition as the validation dataset and the remaining partitions as the derivation dataset. This process will be repeated *k* times and the *k* validation results will be averaged as the final validation result.

Limitations

This study has a couple of limitations. First, due to the strict inclusion criteria, only 13 articles were selected into the final literature review and 15/34 of the highly generalizable risk factors were reported in only one study. Because of the small sample size, it is infeasible to conduct statistical significance tests. However, the intent of this study was not to review risk factors that shared by most studies. Instead, the objective was to provide a list of highly generalizable risk factors to guide the selection of baseline predictor variables in different readmission studies. Even if some risk factors were reported by only one study, we chose to keep them in the list because they were reported to be statistically significant in the prediction of readmissions and can be easily applied to other studies.

Second, the articles are imbalanced in study regions with about half (7/13) based in the United States. This may introduce bias and potentially weaken the generalizability of some risk factors. However, after comparing the US and non-US studies, we did not find an apparent difference in most risk factor categories. Studies in the United States are more likely to use composite comorbidity measures such as comorbidity indices and the number of comorbidities other than individual comorbidities. Although the reported comorbidity indices were originally developed in the United States based on ICD-9-CM codes, they have been translated to work with ICD-10 codes and have been applied globally.

Conclusion

In this work, we have identified 34 highly generalizable risk factors for unplanned 30-day all-cause hospital readmissions. They are not specific to any populations or places and the corresponding predictor

variables can potentially serve as baseline predictor variables in readmission prediction studies around the world. The majority of the identified risk factors are patient-side factors and clinical factors. Only two hospital-side factors have been identified. This could be due to the limitation of the study design and the difficulty of data collection. No major difference has been observed between the risk factors identified inside and outside the United States except that US studies appeared to prefer composite comorbidity measures. However, this assertion should be validated by significance tests when more eligible studies become available. All the reviewed studies have used traditional statistical regression-based methods to identify risk factors. More applications of modern data mining techniques in readmission prediction studies are expected. Because the reviewed studies only explored the association not causation between different variables and readmissions, the identified risk factors should be only used for the predictive modelling of readmissions not for clinical purposes. When more eligible studies become available, this review will be updated by extending the list of highly generalizable risk factors and incorporating statistical analysis to study the variances in different studies. Overall, the literature suggests a growing interest in developing hospital readmission models in the past decade. The findings of this review can guide the selection of baseline readmission predictor variables and potentially provide the foundation for international collaborations on readmission predictions.

Conflicts of Interest

None declared.

References

1. Hospital Quality Initiative - Outcome Measures 2016 Chartbook. <https://web.archive.org/web/20170913132746/https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/OutcomeMeasures.html>
2. Jencks SF, Williams M V., Coleman EA. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. *New England Journal of Medicine* 2009;360(14):1418–1428. PMID: 19339721
3. HSCIC. Hospital Episode Statistics, Emergency readmissions to hospital within 28 days of discharge - Financial year 2011/12. 2013.
4. Kristensen SR, Bech M, Quentin W. A roadmap for comparing readmission policies with application to Denmark, England, Germany and the United States. *Health Policy Elsevier Ireland Ltd*; 2015;119(3):264–273. PMID: 25547401
5. <https://web.archive.org/web/20170913141345/https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html>
6. Boccuti C, Casillas G. Aiming for Fewer Hospital U-turns : The Medicare Hospital Readmission Reduction Program. *Policy Brief*. 2017;1–10.
7. Hines AL, Barrett ML, Jiang J, Steiner CA. Conditions With the Largest Number of Adult Hospital Readmissions by Payer.. *Healthcare cost and Utilization Project 2014*;363(7):1–10. PMID: 24901179
8. Lavenberg JG, Leas B, Umscheid CA, Williams K, Goldmann DR, Kripalani S. Assessing preventability in the quest to reduce hospital readmissions. *Journal of Hospital Medicine* 2014;9(9):598–603. PMID: 24961204
9. Tanzer MW, Heil EM. Why majority of readmission risk assessment tools fail in practice. *Proceedings - 2013 IEEE International Conference on Healthcare Informatics, ICHI 2013* 2013;567–569.
10. Rosen AK, Chen Q, Shin MH, O'Brien W, Shwartz M, Mull HJ, et al. Medical and surgical readmissions in the Veterans Health Administration: what proportion are related to the index hospitalization? *Medical care* 2014;52(3):243–249. PMID: 24374424
11. van Walraven C, Bennett C, Jennings A, Austin PC, Forster AJ. Proportion of hospital readmissions deemed avoidable: a systematic review. *Canadian Medical Association Journal* 2011 Apr 19;183(7):E391–E402. PMID: 25517846
12. Benhassat J, Taragin M. Hospital Readmissions as a Measure of Quality of Health Care. *Arch Intern Med* 2000;160:1074–1081.
13. Alyahya MS, Hijazi HH, Alshraideh HA, Al-Nasser AD. Using decision trees to explore the association between the length of stay and potentially avoidable readmissions: A retrospective cohort study. *Informatics for Health and Social Care* 2017;5498:1–17.
14. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA* 2011 Oct 19;306(15):1688–98. PMID: 22009101
15. <http://www.webcitation.org/6tVYdPwwY>
16. <http://www.webcitation.org/6tVX0hPTg>
17. Dharmarajan K, Hsieh AF, Kulkarni VT, Lin Z, Ross JS, Horwitz LI, et al. Trajectories of risk after hospitalization for heart failure, acute myocardial infarction, or pneumonia: retrospective cohort study. *BMJ (Clinical research ed)* 2015;350(feb05_19):h411. PMID: 25656852
18. 2017 All-Cause Hospital-Wide Measure Updates and Specifications Report: Hospital-Level 30-Day Risk-Standardized Readmission Measure – Version 6.0 <https://web.archive.org/web/20170913132750/https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html>
19. Magdelijns FJH, Schepers L, Pijpers E, Stehouwer CDA, Stassen PM. Unplanned readmissions in younger and older adult patients: the role of healthcare-related adverse events. *European Journal of Medical Research BioMed Central*; 2016 Dec 15;21(1):35.
20. Vest JR, Gamm LD, Oxford BA, Gonzalez MI, Slawson KM, Jencks S, et al. Determinants of preventable readmissions in the United States: a systematic review. *Implementation Science* 2010;5(1):88.
21. 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 Jun 27;6(6):e011060. PMID: 27354072
22. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews* 2015;4(1):1.
23. Berry JG, Zinili SI, Freeman L, Kaplan W, Antonelli R, Gay J, et al. Hospital readmission and parent perceptions of their child's hospital discharge. *International Journal for Quality in Health Care* 2013;25(5):573–581. PMID: 23962990
24. Donzé J, Aujesky D, Williams D, Schnipper JL. Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients. *JAMA Internal Medicine* 2013 Apr 22;173(8):632. PMID: 23529115
25. Van Walraven C, Dhalla IA, Bell C, Etchells E, Stiell IG, Zarnke K, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *Cmaj* 2010;182(6):551–557. PMID: 20194559
26. Tonkikh O, Shadmi E, Flaks-Manov N, Hoshen M, Balicer RD, Zisberg A. Functional status before and during acute hospitalization and readmission risk identification. *Journal of Hospital Medicine* 2016;11(9):636–641. PMID: 27130176
27. Mitchell SE, Sadikova E, Jack BW, Paasche-Orlow MK. Health literacy and 30-day postdischarge hospital utilization. *Journal of health communication* 2012;17 Suppl 3(February 2017):325–338. PMID: 23030580
28. Chakraborty H, Axon RN, Brittingham J, Lyons GR, Cole L, Turley CB. Differences in Hospital Readmission Risk across All Payer Groups in South Carolina. *Health Services Research* 2017;52(3):1040–1060. PMID: 27678196
29. Lin K-P, Chen P-C, Huang L-Y, Mao H-C, Chan D-CD. Predicting Inpatient Readmission and Outpatient Admission in Elderly: A Population-Based Cohort Study. *Medicine* 2016;95(16):e3484. PMID: 27100455
30. Horkan CM, Purtle SW, Mendu ML, Moromizato T, Gibbons FK, Christopher KB. The Association of Acute Kidney Injury in the Critically Ill and Postdischarge Outcomes: A Cohort Study*. *Critical Care Medicine* 2015;43(2):354–364. PMID: 25474534
31. Tan SY, Low LL, Yang Y, Lee KH. Applicability of a previously validated readmission predictive index in medical patients in Singapore: a retrospective study. *BMC Health Services Research* 2013;13(1):366. PMID: 24074454
32. Lin RJ, Evans AT, Chused AE, Unterbrink ME. Anemia in general medical inpatients prolongs length of stay and increases 30-day unplanned readmission rate. *Southern medical journal* 2013;106(5):316–320. PMID: 23644640

33. Bisharat N, Handler C, Schwartz N. Readmissions to medical wards: Analysis of demographic and socio-medical factors. *European Journal of Internal Medicine European Federation of Internal Medicine.*; 2012;23(5):457–460. PMID: 22726376
34. Low LL, Liu N, Wang S, Thumboo J, Ong MEH, Lee KH. Predicting 30-day readmissions in an Asian population: Building a predictive model by incorporating markers of hospitalization severity. *PLoS ONE* 2016;11(12):1–16. PMID: 27936053
35. Blom MC, Erwander K, Gustafsson L, Landin-Olsson M, Jonsson F, Ivarsson K. The probability of readmission within 30 days of hospital discharge is positively associated with inpatient bed occupancy at discharge—a retrospective cohort study. *BMC emergency medicine BMC Emergency Medicine*; 2015;15:37. PMID: 26666221
36. Weeks WB, Lee RE, Wallace AE, West AN, Bagian JP. Do older rural and Urban veterans experience different rates of unplanned readmission to VA and Non-VA hospitals. *Journal of Rural Health* 2009;25(1):62–69. PMID: 19166563
37. Kim H, Hung WW, Paik MC, Ross JS, Zhao Z, Kim GS, et al. Predictors and outcomes of unplanned readmission to a different hospital. *International Journal for Quality in Health Care* 2015;27(6):513–519. PMID: 26472739
38. <http://www.webcitation.org/6tVX0iKPw>
39. Fetter RB, Shin Y, Freeman JL, Averill RF, Thompson JD. Case mix definition by diagnosis-related groups. *Medical care* 1980 Feb;18(2 Suppl):iii, 1–53. PMID: 7188781
40. Halfon P, Eggl Y, Van Melle G, Chevalier J, Wasserfallen JB, Burnand B. Measuring potentially avoidable hospital readmissions. *Journal of Clinical Epidemiology* 2002;55(6):573–587. PMID: 12063099
41. Hasegawa K, Gibo K, Tsugawa Y, Shimada YJ, Camargo CA. Age-Related Differences in the Rate, Timing, and Diagnosis of 30-Day Readmissions in Hospitalized Adults With Asthma Exacerbation. *Chest* 2016 Apr;149(4):1021–9. PMID: 26836926
42. Dreyer RP, Ranasinghe I, Wang Y, Dharmarajan K, Murugiah K, Nuti S V., et al. Sex Differences in the Rate, Timing, and Principal Diagnoses of 30-Day Readmissions in Younger Patients with Acute Myocardial Infarction. *Circulation* 2015 Jul 21;132(3):158–166.
43. Woz S, Mitchell S, Hesko C, Paasche-Orlow M, Greenwald J, Chetty VK, et al. Gender as risk factor for 30 days post-discharge hospital utilisation: a secondary data analysis. *BMJ Open* 2012 Apr 18;2(2):e000428.
44. Rieke K, McGeary C, Schmid KK, Watanabe-Galloway S. Risk Factors for Inpatient Psychiatric Readmission: Are There Gender Differences? *Community Mental Health Journal* 2016 Aug 25;52(6):675–682.
45. Kwan LY, Stratton K, Steinwachs DM, Programs MP, Practice H, Division M. Accounting for Social Risk Factors in Medicare Payment. 2016. PMID: 26844313
46. Baker EH. Socioeconomic Status, Definition. *The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society* Chichester, UK: John Wiley & Sons, Ltd; 2014. p. 2210–2214.
47. Adler NE, Ostrove JM. Socioeconomic Status and Health: What We Know and What We Don't. *Annals New York Academy of Sciences* 1999;896(1):3–15. PMID: 10681884
48. Mackenbach JP, Stirbu I, Roskam A-JR, Schaap MM, Menvielle G, Leinsalu M, et al. Socioeconomic Inequalities in Health in 22 European Countries. *INSERM Unité* 2008;23(687).
49. Kind AJH, Jencks S, Brock J, Yu M, Bartels C, Ehlenbach W, et al. Neighborhood Socioeconomic Disadvantage and 30-Day Rehospitalization. *Annals of Internal Medicine* 2014 Dec 2;161(11):765.
50. Adler NE, Newman K. Socioeconomic Disparities In Health: Pathways And Policies Inequality. *Health Affairs* 2002;21(2):60–76.
51. <https://web.archive.org/web/20170913130757/https://www.medicare.gov/>
52. Joynt KE, Jha AK. A Path Forward on Medicare Readmissions. *New England Journal of Medicine* 2013 Mar 28;368(13):1175–1177. PMID: 23465068
53. <http://www.webcitation.org/6tVXgpuVz>
54. Makowsky MD, Klein EY. Identifying the relationship between length of hospital stay and the probability of readmission. *Applied Economics Letters* 2017 May 9;4(10):1–6.
55. Kalra AD, Fisher RS, Axelrod P. Decreased Length of Stay and Cumulative Hospitalized Days Despite Increased Patient Admissions and Readmissions in an Area of Urban Poverty. *Journal of General Internal Medicine* 2010 Sep 29;25(9):930–935.
56. Ritchie C. Health care quality and multimorbidity: the jury is still out. *Medical care* 2007 Jun;45(6):477–9. PMID: 17515773
57. Fortin M, Soubhi H, Hudon C, Bayliss EA, Akker M v. d. Multimorbidity's many challenges. *BMJ* 2007 May 19;334(7602):1016–1017.
58. Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Archives of internal medicine* 2002 Nov 11;162(20):2269–76. PMID: 12418941
59. Valderas JM, Starfield B, Sibbald B, Salisbury C, Rloand M. Defining comorbidity: implications for understanding health and health services. *Annals Of Family Medicine* 2009;7:357–363. PMID: 19597174
60. Donzé J, Lipsitz S, Bates DW, Schnipper JL. Causes and patterns of readmissions in patients with common comorbidities: retrospective cohort study. *BMJ (Clinical research ed)* 2013;347(dec16_4):f7171. PMID: 24342737
61. Libroero J, Peiró S, Ordifiana R. Chronic comorbidity and outcomes of hospital care: Length of stay, mortality, and readmission at 30 and 365 days. *Journal of Clinical Epidemiology* 1999;52(3):171–179. PMID: 10210233
62. Zekry D, Loures Valle BH, Graf C, Michel JP, Gold G, Krause KH, et al. Prospective Comparison of 6 Comorbidity Indices as Predictors of 1-Year Post-Hospital Discharge Institutionalization, Readmission, and Mortality in Elderly Individuals. *Journal of the American Medical Directors Association Elsevier*; 2012;13(3):272–278. PMID: 21450226
63. Charlson ME, Pompei P, Ales KL, MacKenzie R. A new method of classifying prognostic in longitudinal studies: development and validation. *Journal of Chronic Diseases*. 1987. p. 373–383. PMID: 3558716
64. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity Measures for Use with Administrative Data. *Medical care* 1998;36(1):8–27.
65. Sharabiani MTA, Aylin P, Bottle A. Systematic review of comorbidity indices for administrative data. *Medical Care* 2012;50(12):1109–1118. PMID: 22929993
66. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of Clinical Epidemiology* 1992;45(6):613–619. PMID: 1607900
67. Leidy NK. Functional status and the forward progression of merry-go-rounds: toward a coherent analytical framework. *Nursing Research*. 1994. p. 196–202.
68. Greysen SR, Stijacic Cenzer I, Auerbach AD, Covinsky KE. Functional Impairment and Hospital Readmission in Medicare Seniors. *JAMA internal medicine* 2015;175(4):559–565. PMID: 24655651
69. Ratzan SC, Parker RM. Introduction. In: Ratzan SC, Parker RM, Selden CR, Zorn M, editors. *National Library of Medicine Current Bibliographies in Medicine: Health Literacy* National Institutes of Health, U.S. Department of Health and Human Services.; 2000. p. v.
70. Health Literacy. Washington, D.C.: National Academies Press; 2004.
71. Evangelista LS, Rasmussen KD, Laramee AS, Barr J, Ammon SE, Dunbar S, et al. Health Literacy and the Patient With Heart Failure—Implications for Patient Care and Research: A Consensus Statement of the Heart Failure Society of America. *Journal of Cardiac Failure* 2010;16(1):9–16. PMID: 20123313
72. Cloonan P, Wood J, Riley JB. Reducing 30-Day Readmissions: Health Literacy Strategies. *JONA: The Journal of Nursing Administration* 2013;43(7/8):382–387.
73. Clancy C. An Overview of Measures of Health Literacy. *Institute of Medicine (US) Roundtable on Health Literacy Measures of Health Literacy: Workshop Summary* Washington, D.C.: National Academies Press (US); 2009.
74. Encinosa WE, Hellinger FJ. The Impact of Medical Errors on Ninety-Day Costs and Outcomes: An Examination of Surgical Patients. *Health Services Research* 2008 Dec;43(6):2067–2085.
75. <http://www.webcitation.org/6tVXnkj32>
76. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical care* 2005 Nov;43(11):1130–9. PMID: 16224307
77. Halfon P, Eggl Y, van Melle G, Chevalier J, Wasserfallen JB, Burnand B. Measuring potentially avoidable hospital readmissions. *Journal of clinical epidemiology* 2002 Jun;55(6):573–87. PMID: 12063099

-
78. Nelder JA, Wedderburn RWM. Generalized Linear Models. *Journal of the Royal Statistical Society Series A (General)* 1972;135(3):370–384.
79. Cox DR. The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society Series B (Methodological)* 1958;20(2):215–242.
80. CORNFIELD J. A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. *Journal of the National Cancer Institute* 1951 Jun;11(6):1269–75. PMID: 14861651
81. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang J-F, et al. Data mining in healthcare and biomedicine: a survey of the literature. *Journal of medical systems* 2012 Aug;36(4):2431–48. PMID: 21537851
82. Quinlan JR. *Discovering Rules by Induction from Large Collections of Examples. Expert Systems in the Micro Electronic Age* Edinburgh University Press; 1979.
83. Piatetsky-Shapiro G. Discovery, analysis, and presentation of strong rules. *Knowledge Discovery in Databases AAAI/MIT Press*; 1991. p. 229–248.
84. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *European Urology* 2015;67(6):1142–1151. PMID: 25561516
85. Krumholz HM, Brindis RG, Brush JE, Cohen DJ, Epstein AJ, Furie K, et al. Standards for statistical models used for public reporting of health outcomes: An American Heart Association scientific statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group. *Circulation* 2006;113(3):456–462. PMID: 16365198
86. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983 Sep;148(3):839–843.
87. Green DM, Swets JA. *Signal detection theory and psychophysics*. New York, New York, USA: John Wiley & Sons, Inc; 1966.
88. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2* 1995. p. 1137–1143.