

A Roadmap for Designing a Personalized Search Tool for Individual Healthcare Providers

Gang Luo

Department of Biomedical Informatics, University of Utah
HSEB Room 5725B, 26 South 2000 East, Salt Lake City, UT 84112, USA
gang.luo@utah.edu

Abstract

Each year, a large percentage of people change their physicians and other individual healthcare providers (IHPs). Many of these people have difficulty identifying a replacement they like. To help people find satisfactory IHPs who are likely to be good at managing their health issues and serve their needs well, in a previous paper we proposed a high-level framework for building a personalized search tool for IHPs. There are many issues regarding designing a personalized search tool for IHPs, of which only a small portion are mentioned in our previous paper. This paper surveys various such issues that are not covered in our previous paper. We include some preliminary thoughts on how to address these issues with the hope to stimulate future research work on the new topic of personalized search for IHPs.

Keywords: Personalized search for individual healthcare providers · individual healthcare provider effect · predicting healthcare cost · predicting care quality measures · predicting patient satisfaction

1. Introduction

Consumers frequently change their individual healthcare providers (IHPs) such as physicians, dentists, and dietitians. As reported by RAND Health [1], 38% of Americans changed physicians within the past two years. Of those people who changed physicians, 36% encountered difficulty identifying a replacement they like. For people in fair or poor health, this percentage increases to 55%. Since patient satisfaction, health outcomes, and healthcare costs are all affected by the patients' selection of IHPs, this phenomenon is undesirable.

The degree of satisfaction of a patient with his IHP correlates with his health outcome. A patient who is more satisfied with his IHP is more likely to listen to his IHP's advice and realize various health benefits including medication compliance, adopting a healthy lifestyle, and complying with preventive measures [2, 3, 4].

Similar to the way therapist effect is defined in Baldwin and Imel [5], we define *IHP effect* as the effect of an IHP on a specific patient's care quality, healthcare cost, and degree of satisfaction as compared to another IHP. As known by many clinicians making referrals to colleagues, IHP effect can be large [5, 6, 7, 90] partly because differing

IHPs often treat the same condition in rather different ways [8, page 4, 9, pages 74 and 396]. Even with expert training and close supervision, IHP effect will not become minimal [10]. To optimize health system performance, the Institute for Healthcare Improvement developed the Triple Aim to improve patient health, decrease healthcare cost, and increase patient satisfaction [11]. A large part of the current healthcare crisis comes from skyrocketing healthcare costs. Under a capitated payment model in which a healthcare provider receives a fixed annual amount per health insurance plan member, reducing healthcare cost can both benefit society and increase the healthcare provider's profit margin. Choosing an effective IHP for each patient is important for achieving the Triple Aim.

To help people find satisfactory IHPs who are likely to be good at managing their health issues and serve their needs well [9, page 347], we previously proposed a high-level framework for building a personalized search tool for IHPs [12]. The framework combines the use of both health knowledge and patients' medical information with the current trend of personalized search [13] and contextual search [14] in Web search. As shown in Fig. 1, the search tool can be implemented as a new function in a personal health record system connecting to the provider's electronic medical record system and administrative system. Alternatively, the search tool can be implemented as a stand-alone online tool that has access to information extracted from the provider's electronic medical record system and administrative system. In the latter case, the user needs to input more information describing his needs than that in the former case.

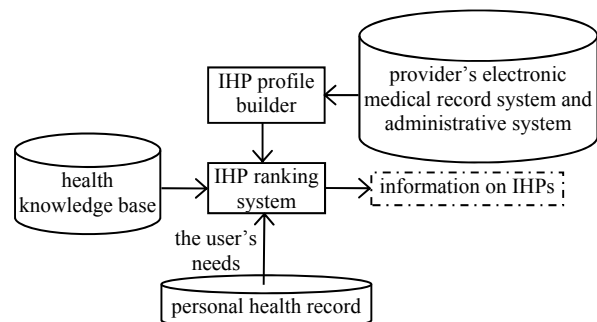


Fig. 1 Workflow of the personalized search tool for IHPs.

Our high-level idea of building the personalized search tool for IHPs is to use health knowledge, and introduce and extend data mining technology [15] and recommender system technology [16] into the personal health record domain. To obtain more information on an IHP and expand her profile, the historical data of all of her patients is first extracted from the provider's electronic medical record system and administrative system, and then aggregated as the extended component of her profile. The aggregated information reflects the care quality, healthcare costs, and patient satisfaction when the IHP manages patients with various health issues. The needs of the user are obtained from the information stored in his personal health record, the priorities specified by him, his desired type of IHP, and his inputted preferences. Then using health knowledge stored in the health knowledge base, all available IHPs of the user's desired type are sorted according to how well their profiles match his needs.

To perform sorting, we predict for each such IHP the user's care quality measures, healthcare cost, and degree of satisfaction with the IHP if the user is going to be managed by the IHP. The higher-ranked IHPs are more likely to satisfy the user, help him achieve a good health outcome, and/or keep his healthcare cost as low as possible. For each displayed IHP, multiple attributes, such as performance information, in her profile are presented to help the user understand the search results [17] and make an informed decision.

There are many issues regarding designing a personalized search tool for IHPs, of which only a small portion are mentioned in our previous paper [12]. To fill the gap, this paper surveys various such issues that are not covered in our previous paper. We include some preliminary thoughts on how to address these issues with the hope to stimulate future research work on the new topic of personalized search for IHPs.

The rest of the paper is organized as follows. Section 2 provides some background on IHP effects. Section 3 describes our expectations of the personalized search tool for IHPs. Section 4 discusses several generic issues on personalized search for IHPs. Section 5 presents issues specifically related to predicting healthcare costs. Section 6 addresses issues specifically related to predicting care quality measures. Section 7 mentions issues specifically related to predicting patient satisfaction. Section 8 concludes this paper.

2. Background on IHP effects

In this section, we present some background on IHP effects. The personalized search tool for IHPs has enough room to optimize the matching between patients and IHPs because IHP effects are often large.

2.1 Impact of IHP effects

Physicians make more than 80% of the decisions on various healthcare issues, such as whether hospitalization is

needed, the length of hospital stay, the (types of) medications ordered, the diagnostic tests used, and the management plans adopted [18]. These decisions have a major impact on patient health outcomes and healthcare costs. Patient health outcomes are the most important factor in determining patient satisfaction degree in the long term [19].

Differing IHPs often treat the same condition in rather different ways [8, page 4, 9, pages 74 and 396]. As a result, the IHP effect can be large [5, 6, 7, 90]. In other words, if the same patient is treated by differing IHPs for the same condition, the patient's care quality, healthcare cost, and/or degree of satisfaction will often be rather different.

For instance, in psychotherapy, about one-third of the treatment effect comes from therapist effect [5, 10, 20]. As reported in Saxon and Barkham [20], therapist effect is larger for more severe patients. For the most effective therapists, the patient recovery probability is about two times that of the least effective therapists. For the least effective therapists, the patient deterioration rate is about three times that of the other therapists. Brown *et al.* [21] found that patients of effective therapists showed three times as much improvement as the rest of the patients in their study.

One manifestation of IHP effect is the variation in practice patterns within the same hospital. As mentioned in James and Savitz [90], the patients of different IHPs in the same hospital often have similar levels of severity or complexity in their medical conditions. An IHP is found to have a consistent usage rate of healthcare resources. However, usage rate of individual treatment elements varies by 1.6-5.6 times across different IHPs. For each treatment, patient healthcare cost, excluding payments to IHPs, also varies by two times across different IHPs.

Another manifestation of IHP effect is the well-known geographic variation in practice patterns [8, page 4]. For example, Medicare reimbursements vary more than threefold in different regions [22]. Most of the variation in Medicare spending has nothing to do with regional differences in prices, but instead comes from the volume of healthcare services used, which is not justified by patient need for healthcare services [18, 22]. In fact, no improved quality of care, life expectancy, or health outcome is observed in regions with higher Medicare spending [18]. It is estimated that such variation in Medicare spending with no positive effect costs Medicare about 29% of its total expenditure [18]. As a second example, about half of the variation in psychiatric inpatient length of stay and cost results from the variation of practice patterns in different facilities [9, page 397, 23]. As a third example, rates of medication use in treating mood disorders vary fourfold across similar clinics [9, page 396, 24].

2.2 Factors leading to large IHP effects

Several factors cause IHP effects to often be large. One major factor is related to health problems. As explained in

Elstein *et al.* [25, 26, Chapter 4, 90], the scope of healthcare is so large and the amount of health knowledge grows so rapidly that no IHP can comprehensively master every domain of healthcare. No generic strategy can enable an IHP to succeed in handling every health problem. Instead, domain-specific knowledge plays a major role in an IHP's success in dealing with a specific health problem. Thus, differing IHPs are good at managing different health problems. Also, the performance of the same IHP varies widely for different health problems [26, Chapter 4, 27].

For chronic conditions, IHP effect will accumulate over time and have a substantial impact on the patient's functioning in the long run, even if the effect is small in a single year [5]. According to the Centers for Disease Control and Prevention [28], about 50% of American adults have chronic conditions. Over 75% of healthcare costs are due to chronic conditions. By considering only chronic conditions, IHP effects can already have a huge impact on public health.

Besides health problems, personal characteristics also play a role in IHP effects. For example, a therapist can be more effective at managing certain types of patients (e.g., Causasian or racial/ethnic minority) than others [29]. As a second example, for certain health issues, the age and specialty of the attending physician can help predict the variation in healthcare service usage [23]. To use IHP effects well to our advantage, both health problems and personal characteristics need to be considered.

2.3 Existing efforts on reducing the likelihood of choosing ineffective IHPs

Health insurance plans have long realized that IHP effect can be large. To reduce the likelihood of choosing ineffective IHPs, health insurance plans offer patients differential copayments to encourage them to visit physicians providing higher-quality and lower-cost services, pay bonuses to physicians whose use of resources is lower than average, and publicly report the relative costs of physicians' services [30]. In particular, the Centers for Medicare & Medicaid Services publicly reports physician-level quality measure data on the Physician Compare Web site [31], hospital mortality and readmission data on the Hospital Compare Web site [32], nursing home quality measure data on the Nursing Home Compare Web site [33], and home health agency quality measure data on the Home Health Compare Web site [34].

3. Our expectations of the personalized search tool for IHPs

In this section, we describe our expectations of the personalized search tool for IHPs. These expectations will affect the criteria according to which we optimize the search tool's performance.

3.1 The need for and the usage of the personalized search tool for IHPs

Consumers often have difficulty determining the right IHPs for them. Many patients are bad at judging the technical quality of the care they receive and allow care to be of extremely poor quality before expressing dissatisfaction [19, 35]. They have no idea what the best care for their health issues is, and hence cannot tell how well their IHPs are managing their health issues. Frequently, professional guidance is unavailable to consumers when they are seeking the right IHPs for their specific needs. This is typically the case when consumers need to change their IHPs due to relocation or insurance change. Even if consumers have access to clinicians to request recommendations, many clinicians can provide only a limited amount of useful information because they know little about the expertise of their colleagues [25]. Also, IHPs usually have no way of comparing their treatment outcomes with those obtained by other IHPs [36], and thus cannot independently determine whether they are the right IHPs for their patients.

It would be desirable to build a personalized search tool for IHPs to assist consumers in finding the right IHPs [9, page 347]. This would reduce the likelihood that IHPs run into the types of patients that they are bad at managing, and help IHPs focus on the types of patients that they are good at managing [9, page 347]. The search tool uses a data mining process [12], which can help us better understand why an IHP works well with some patients but not with others [5]. If we can identify a particular type of patient with whom the IHP works poorly as well as why, then if necessary, we may give the IHP more supervision or further training to help her improve her skill of managing this type of patient [37, 38].

3.2 Our expectations with rationale

Based on our understanding of various properties of IHP effects, we have certain expectations of the personalized search tool for IHPs. Although IHP effect is often large, it can be minimal. In many cases, a consumer will have a similar health outcome, healthcare cost, and/or degree of satisfaction regardless of which IHP he sees [5]. We call such a consumer insensitive. Many relatively healthy consumers without major health issues are insensitive. For an insensitive consumer, how the search tool ranks IHPs matters little. What matters most in practice is how the search tool ranks IHPs for the sensitive consumers on whom IHP effects can be potentially large. Therefore, we should focus on optimizing the search tool's performance for the sensitive consumers.

For a given patient, it is difficult, if not impossible, to predict very precisely which IHP will be the best one for him to provide the best health outcome, the lowest healthcare cost, and/or the highest degree of satisfaction. In many cases, a large portion of IHPs are effective for the patient [20]. Hence, the goal of the personalized search tool for IHPs is not to find the exact best single IHP for each user. Rather, the goal is to help the user avoid choosing those IHPs

that will result in a much worse health outcome, much higher healthcare cost, or much lower degree of satisfaction than what the most effective IHPs for him will bring [38]. That is, the search tool should rank the most ineffective IHPs low, particularly for the sensitive users [39].

The top few IHPs ranked by the search tool should be effective for the user, whereas it does not matter much if they are not the most effective ones for the user. It is acceptable if some effective IHPs for the user are ranked low by the search tool [39]. Usually, multiple IHPs are effective for a user [20]. There is no need to identify all effective IHPs. If the search tool can identify some of them and subsequently the user chooses one of the identified, effective IHPs, it is good enough. In general, for a sensitive user, the potential negative impact of ranking a very ineffective IHP high is much larger than that of ranking a very effective IHP low.

As a manifestation of the 80-20 rule that holds in many places, a small portion of the population consumes most of the healthcare costs [9, page 254]. 2% of the patients consume about 40% of the healthcare costs [40]. 10% of the patients consume about two thirds of the healthcare costs [41]. Many of these patients are sensitive. Thus, we would expect to achieve significant benefit just by eliminating the possibility of choosing the most ineffective IHPs for the sensitive patients, although these cases represent only a small subset of all possible cases for selecting IHPs for patients. This is consistent with the observation that most of the potential for healthcare cost savings is concentrated in a small fraction of cases [9, page 270].

Due to random variation, even if a patient sees the worst/best IHP for him, there is still a decent chance that he will have an above/below average outcome [5, 38], particularly in the short term. Thus, regardless of how good the personalized search tool for IHPs is, its use cannot guarantee achieving the best possible result on any single user. We can only expect that on average its use will improve the results for all users.

The personalized search tool for IHPs is used to help a patient find the right IHP for him. As mentioned in Duncan [8, page 259], if a patient finds the right IHP for him, his physician, lab test, and drug costs may increase due to better compliance with ordered treatments. In contrast, his total healthcare cost is likely to decrease as a result of a shortened hospital stay, reduced number of hospital admissions and other high-cost, potentially life-threatening events, particularly in the long term. In general, a high-cost medical event is much more expensive than an ordinary healthcare event. For instance, an ambulatory visit costs about \$150 on average, while an inpatient stay costs about \$6,000 on average [42]. By avoiding a single high-cost medical event, we can immediately realize financial benefit.

As mentioned in [18], the public reporting of hospital and IHP performance is a powerful driver of hospital and IHP performance improvement. Even the threat of disclosing

performance information to the public can significantly increase quality improvement activities and improve performance. Since the personalized search tool for IHPs displays IHP performance information [12], we would expect the search tool to improve IHP performance.

Usually, healthcare data is imperfect and has various issues with its accuracy, size, and completeness. Due to the healthcare crisis, healthcare providers have to figure out with great urgency how to improve care delivery based on imperfect data [9, page 448]. The personalized search tool for IHPs fits with this healthcare industry trend. Although the search tool will be imperfect for various reasons, having it outweighs the alternative of having no tool.

4. Generic issues on personalized search for IHPs

In this section, we discuss several generic issues on personalized search for IHPs.

4.1 Conflicts among multiple criteria

The degree of matching between an IHP's profile and the user's needs is computed using a fixed set of criteria, such as the user's care quality measures, degree of satisfaction with the IHP, and healthcare cost. These criteria frequently conflict with each other [9, page 198] and are not equally important. In case of conflict, we should emphasize the few criteria that are regarded as the most important ones according to health knowledge, while striving to avoid extremely undesirable values with the other criteria. This can be done using multi-criteria recommender system techniques [16].

For example, patients often request discretionary services that have little or no medical benefit, but may lead to iatrogenic harm as a result of overtreatment or other causal pathways [43]. To provide quality care and reduce healthcare costs, physicians should refuse these requests. This can cause lower patient satisfaction that is not associated with any problem in care quality [4, 43]. In this case, the criterion of patient satisfaction conflicts with the criteria of patient care quality and healthcare cost. As mentioned in Jackson *et al.* [19, 35], patient satisfaction often does not correlate well with patient care quality because many patients are bad at assessing care quality. For instance, patient satisfaction is higher when tests are ordered, even if the ordered tests have no diagnostic value.

Typically, patient care quality is more important than patient healthcare cost, which is subsequently more important than patient satisfaction. Without using additional measures to ensure care quality, we may erroneously favor those IHPs who achieve low healthcare costs by withholding necessary care [44]. At the same time, an overemphasis on patient satisfaction can lead to adverse consequences on healthcare resource utilization, costs, and outcomes [43]. Thus, by default we would give more emphasis to patient care quality than to healthcare cost, and more emphasis to healthcare cost than to patient satisfaction. If the user really wants, he has the option to

change the criteria's priorities in the personalized search tool for IHPs [12].

4.2 Time frames on which criteria are based

Each criterion of patient healthcare cost, care quality measures, and satisfaction is specified based on a time frame, such as the patient's total healthcare cost within the next 12 months [9, page 18]. Often, the value of a criterion depends a lot on the length of its corresponding time frame resulting in a complex relationship between them.

For example, as noticed by Jackson *et al.* [19], patient satisfaction in the short term is often different from that in the long term. The way the IHP communicates with the patient strongly correlates with patient satisfaction in the short term, such as immediately after the patient's visit to the IHP. The correlation diminishes quickly over time. The patient's health outcome as a result of the IHP's treatment eventually becomes the most important factor in determining patient satisfaction in the long term, such as two weeks or longer after the patient's visit to the IHP.

As a second example, a patient with a chronic health issue may incur roughly the same amount of healthcare cost every month. In contrast, another patient recently diagnosed with an acute health issue will go through some treatment in the next few weeks. After the acute health issue is cured, he is likely not to incur any healthcare cost in the next few months. In this case, if we predict the patient's total healthcare cost within the next month and multiple it by 12, we will not obtain a good estimate of the patient's total healthcare cost within the next 12 months. To predict a patient's total healthcare cost within a time frame more accurately, we need to build a prediction model specific to the time frame.

Sometimes, if a criterion is based on a time frame that is not long enough, we cannot use the criterion to show the performance differences among different IHPs. For instance, as mentioned in Iezzoni [9, page 397], psychiatric disorders are mainly chronic conditions. Psychosocial functioning generally improves late in the course of treating a psychiatric disorder. In this case, a long enough time frame is needed to see the treatment effect.

In our application of matching patients with IHPs, we generally focus more on long-term results than on short-term results because long-term results are what eventually matter. Hence, each criterion of patient healthcare cost, care quality measures, and satisfaction usually should be based on a time frame that is long enough. For certain criteria such as patient healthcare cost, the corresponding time frame will be an integral number of years to guard against seasonal trends in healthcare service usage [45].

4.3 Balancing the IHPs' workloads

To decrease patient wait time and improve continuity of care, Balasubramanian *et al.* [46] proposed redistributing patients across primary care physicians (PCPs) to balance physicians' case mixes and subsequently workload. This

can also help improve healthcare efficiency [47]. If a patient is forced to see a new PCP because his regular PCP is overloaded, the new PCP will need to invest time to get familiar with his medical history. In the personalized search tool for IHPs, it would be beneficial to consider the IHPs' case mixes in attempt to reduce the likelihood of overloading them. For instance, we can put the IHP's case load as one of the criteria used in computing the degree of matching between the IHP's profile and the user's needs.

4.4 A patient managed by multiple IHPs simultaneously

Sometimes, a patient's care quality, healthcare cost, and/or degree of satisfaction depend on the performance of multiple IHPs rather than a single IHP [25, 27, 48]. For example, this is the case with surgery or when a patient sees multiple PCPs simultaneously. The latter often occurs with Medicare patients [8, page 36]. In the case of multiple IHPs, if one IHP acts sub-optimally, the patient can have a poor outcome, a high healthcare cost, and/or low satisfaction. The challenge is that we may not be able to identify which IHP performed sub-optimally, partly because medical practice is an art. There is no evidence-based or absolute standard of what level of resource use is optimal for every possible combination of health issues [18].

In general, when a patient is managed by multiple IHPs simultaneously, it is difficult to determine the exact portion of the patient's care quality, healthcare cost, and/or degree of satisfaction that should be attributed to each IHP. Nevertheless, in comparing the performance of various IHPs, ideally the exact portion should be used.

A traditional method for dealing with situations involving multiple IHPs is to divide the entire span of patient care into episodes and attribute each episode to a single IHP [8, page 265, 30]. An episode of care is "a series of temporally contiguous healthcare services related to treatment of a given spell of illness or provided in response to a specific request by the patient or other relevant entity" [9, page 84, 49]. The software used for identifying healthcare episodes is called episode grouper.

A healthcare episode is typically attributed to the dominating IHP [8, page 265, 30]. For example, a surgical episode is attributed to the surgeon with the highest allowed charge. A non-surgical episode is attributed to the IHP with the largest number of visits, the most recent service, or the most expensive service. Some work specifies a minimum threshold for the degree of dominance [30]. For instance, at least a certain percentage, such as 30%, of the healthcare cost in an episode should be billed to the IHP to whom the episode is attributed. If no such IHP can be found, the episode will not be attributed to any IHP.

Frequently, a healthcare episode related to a chronic condition is divided into phases, such as a diagnostic phase (when the chronic condition is first recognized), a maintenance phase (routine management), an acute flare-up phase, and a chronic complication phase. Each phase can have its own set of risk factors [9, page 87]. If needed, we

can build a separate set of prediction models for each phase to predict the patient's healthcare cost, care quality measure, and degree of satisfaction associated with the phase. The predicted results for all of the phases in an episode are merged together to become the predicted results for the episode. The predicted results for one or more episodes are merged together to become the predicted results for the patient.

The episode method may not work well in certain cases. We are unaware of any other method that is clearly better than the episode method for dealing with the situation of having multiple IHPs. Sometimes, the boundaries between episodes of care are fuzzy. For example, this is the case with many people, especially elderly patients, who have multiple concurrent chronic conditions that wax and wane [9, page 85]. Also, we do not fully understand the performance of existing commercial episode groupers because they have not been thoroughly tested for either validity or agreement with each other [18].

4.5 Incomplete medical data on a patient resulting from the use of multiple IHPs from different provider groups

One aim of the personalized search tool for IHPs is to help reduce each patient's total healthcare cost. To the extent possible, we would like to capture all healthcare costs spent on each patient, regardless of who paid for it. Typically, healthcare expenditures include inpatient stays, outpatient visits, emergency department visits, prescription medications, home health visits, dental visits, and medical supplies including vision and hearing aids [41, 50]. Although they can be substantial, payments for over-the-counter drugs and most alternative care are not included because of the difficulty of keeping track of them [51]. For example, for a patient with chronic conditions, payments for health-related services often reach thousands of dollars per year [9, page 96]. If a patient has multiple health insurances, his total healthcare cost is the sum of the healthcare costs paid by each of these insurances, his out-of-pocket payments, and any unpaid part, e.g., due to denied claims [42].

As mentioned in Section 1, the personalized search tool for IHPs uses the patients' medical data to build the IHPs' profiles. Thus, the search tool's performance depends on the portion of each patient's medical data that is available to the search tool. The ideal situation is that the provider's electronic medical record system and administrative system have all of the medical data collected on the patient. This is usually the case when a patient is managed by a single IHP. For instance, a patient with a severe health issue often prefers to see a single IHP who is familiar with his medical history and/or experienced at managing his health issue, regardless of the IHP's office location [52].

When a patient is managed by multiple IHPs simultaneously and these IHPs belong to the same provider group, the provider's electronic medical record system and administrative system usually have all of the medical data

collected on the patient. For example, this is typically the case when the patient uses an HMO (Health Maintenance Organization) plan as his sole health insurance plan [53]. As a second example, an American veteran with low income is likely to use only Veterans Affairs (VA) hospitals because of the low premiums of VA health insurance plans.

When a patient is treated by multiple IHPs simultaneously and these IHPs belong to different provider groups, a missing data problem occurs. In this case, no single provider group has complete information on the patient, such as all of his healthcare costs and lab test results. The payer of claims, such as the health insurance plan, often has more complete data than the individual provider groups [8, page 36]. In case it is difficult to obtain data from multiple provider groups, it can be beneficial to combine the data from a provider group with the data from health insurance plans, such as the data in an all-payer claims database [54] or the data from Medicare and Medicaid [55]. If data from multiple sources are coded based on different clinical coding systems, we need to map the codes in these systems into the same set of codes before the data can be combined. Such mappings are routinely used in the enterprise data warehouses of multiple provider groups.

As shown in considerable work [8, 9, 56, 57, 58], medical claims data alone provides much useful information for various prediction purposes. The most commonly used data sources include the provider's electronic medical record system, medical claims, and health insurance enrollment files containing basic demographic and health insurance eligibility information [9, page 97]. Electronic medical record data provides various details about the clinical complexity of a patient's health issues. Many of these details are not covered by coded data in medical claims [9, page 134, 59].

In addition to these data sources, another frequently used data source is death certificate information from the Department of Health of state governments [60]. Deaths that occurred outside of healthcare organizations are captured in this data source, but missed in many other data sources. In general, the more data sources we use, the more precisely we can predict patient healthcare cost, care quality measures, and satisfaction [9, pages 84 and 282, 51, 59, 61, 62, 63]. These predicted values are used in computing the degree of matching between an IHP's profile and the user's needs [12].

4.6 Users with varying amounts of available information

In the personalized search tool for IHPs, all prediction models are trained using data from one or more sources. One data source is the provider's electronic medical record system. Each patient appearing in the training data typically has many attributes recorded in the provider's electronic medical record system. In contrast, the information about a user of the search tool comes from his inputs and/or his personal health record. The amount of this information can vary significantly from one user to another [51]. For example,

a new user may input only the most basic information into the search tool and have an empty personal health record. Another long-term user may input a lot of information into the search tool and/or already have ample information accumulated in his personal health record.

In general, the more information we have about the user, the better the personalized search tool for IHPs performs. The search tool can prompt for information about the user, but should not force the user to input a lot of information. For the search tool to have good usability, the user should have the flexibility to decide how much information he is willing to input.

Consider a user and a prediction model. Each user attribute whose value is missing but needed by the prediction model is assigned a default value [12]. If many default attribute values are assigned, we can easily make the information about the user less reliable and cause the prediction model to perform poorly.

To use each user's available information to the fullest, we can build multiple sets of prediction models in the personalized search tool for IHPs. Each set of prediction models corresponds to a different set of user attributes. To make predictions on a given user, we choose an optimal set of prediction models from the available sets using two criteria. First, the set of prediction models should use as much available information about the user as possible. Second, the amount of missing information needed by the set of prediction models should be minimal. In other words, the set of user attributes used in the optimal set of prediction models should be as close as possible to the set of user attributes with non-empty values.

4.7 Prediction algorithms

Most of the existing work on predicting patient healthcare cost, care quality measures, or satisfaction uses linear statistical models, such as linear regression and logistic regression [8, 9, 19, 42, 59, 64, 65, 66, 67]. The prediction accuracy achieved in such work is often fairly low, partly because some clinical variables have complex interactions among each other [68, 69, 70, 71] while most clinical variables have a nonlinear relationship to patient healthcare cost, care quality measures, and satisfaction [9, page 216, 45]. For instance, in predicting each patient's healthcare cost, the average prediction error is usually as large as the average healthcare cost [42]. We would expect appropriately designed, nonlinear machine learning methods [72] to perform better than linear statistical models in making these predictions.

Healthcare data is often distributed in an uneven way. Each instance belongs to a class. In predicting patient healthcare cost, care quality measures, and satisfaction, we frequently run into the imbalanced data problem, where one or more classes are associated with a much smaller portion of all of the instances than the average. For example, as is the case with healthcare utilization data in general [45, 73], each year only a small portion of people have high

healthcare costs, whereas most people have zero or low healthcare costs [9, page 254, 40, 42, 51, 65]. In the research literature, many techniques have been developed for addressing the imbalanced data problem [72, Chapter 16]. Using these techniques can help improve the prediction accuracy for the minority classes.

As is typical with machine learning, ensemble methods are likely to perform better than individual machine learners [74]. For each criterion of patient healthcare cost, care quality measures, and satisfaction, its value should be predicted using a different machine learning model and set of features [9, pages 197 and 198]. A machine learning model and set of features originally designed for another purpose often predict a criterion's value less precisely than another machine learning model and set of features specifically designed for this purpose [9, pages 17, 84, and 198, 60, 75]. Even if two care quality measures are predicted using the same machine learning method and similar features, the features' weights in the prediction model vary by the care quality measure [9, page 198].

All prediction models used in the personalized search tool for IHPs need to be updated regularly to remain current. As new data keeps arriving, healthcare techniques keep advancing, and medical practice patterns keep changing, the prediction models need to be retrained [9, pages 387 and 389]. Some features also need to be modified periodically, e.g., as new drug codes are released every month [8, pages 54 and 79].

4.8 Performance issue diagnosis on prediction models

For the personalized search tool for IHPs to work well, a certain level of prediction accuracy needs to be achieved on each criterion of patient healthcare cost, care quality measures, and satisfaction. If a prediction model for a criterion cannot achieve our desired level of prediction accuracy, we can conduct a performance issue diagnosis on the prediction model in the following way.

We divide the data into subgroups to identify the types of cases for which the prediction model works poorly [9, page 288]. The subgroups are determined based on some patient properties, such as age, gender, and diagnosis. This division can help us determine why predictions are poor for some cases. Once we know the reason, we can adopt a corresponding makeup strategy to improve the model's prediction accuracy. For instance, we can add into the prediction model features designed specifically for addressing the reason.

4.9 Sample size considerations

A typical PCP practicing full time has about 2,000 patients, i.e., a patient panel size of 2,000 [46]. In building an IHP's profile, we need to consider several issues related to sample size.

First, for an outcome measure assessing a rare event, an IHP's patient panel size can be too small to meaningfully examine the average outcome measure of the IHP's patients.

In this case, we can exclude the outcome measure from the IHP's profile. Alternatively, we can aggregate multiple related outcomes into one super-outcome, either for a single health issue or across several types of care (e.g., preventive, acute, chronic) [9, page 198, 18]. The measure of the super-outcome is a composite one.

Second, for a rare health issue, an IHP's patient panel size can be too small to meaningfully examine the average outcome measures of the IHP's patients with the health issue [5, 9, page 75]. If an IHP has only a small number of patients with a particular health issue, we can exclude from the IHP's profile the average outcome measures specifically related to the health issue. Alternatively, we can set each such average outcome measure in the IHP's profile to a default value, such as the average outcome measure of all of the patients with the health issue.

Performance profiles are often used to judge each IHP's absolute performance. For this purpose, a certain minimum number of patients with a specific health issue are needed for producing an IHP's performance profile. This minimum number is usually relatively large, such as 30 [9, page 358, 18], because the IHP's absolute performance needs to be known precisely. In comparison, in the personalized search tool for IHPs, we also need a certain minimum number of patients to meaningfully compute an average outcome measure of an IHP's patients with a specific health issue. The average outcome measure is used in the IHP's profile to make various predictions. Since the search tool's goal is to help the user avoid bad IHPs for him, we neither need very precise predictions nor have to know each IHP's absolute performance precisely. Thus, the accuracy requirement on the average outcome measure is lower than that on producing an IHP's performance profile. In other words, the minimum number of patients needed by the search tool is smaller than the one needed for judging each IHP's absolute performance.

4.10 Health issue-specific prediction models

Each health issue is different and has its own unique properties. A single prediction model may not work well for all possible health issues. For certain health issues, it can be beneficial to build separate models specific to each health issue to predict patient healthcare cost, care quality measures, or satisfaction [9, page 202, 76]. For example, as mentioned in Ash *et al.* [9, page 376, 42], there is great variability in the healthcare costs of the people in a heterogeneous group, with a coefficient of variation usually between 3 and 4. However, when we examine the healthcare costs of the patients with a specific health issue, the variability often becomes much smaller. In general, the smaller the variability, the easier it is to predict the healthcare costs.

In using health issue-specific prediction models, the following factors need to be considered. First, many patients have multiple health issues. For such a patient, several health

issue-specific prediction models may apply to him. In this case, it is non-trivial to decide which prediction model to use.

Second, in general, a prediction model specific to a health issue should use features related to various health issues rather than features only related to this particular health issue. For instance, the healthcare cost of a patient with a specific health issue is the cost of treating all of his health issues rather than only the cost of treating the specific health issue [42]. Without considering the patient's other health issues, we may significantly underestimate the patient's healthcare cost.

Third, developing and validating a health issue-specific prediction model requires much data [76]. It is possible to obtain enough data for common health issues. However, it is difficult to obtain enough data for rare health issues. Hence, health issue-specific prediction models are feasible mainly for common health issues.

4.11 Health issues of focus

There are many health issues. Due to limited amounts of available data and resources, we can neither build a set of health issue-specific prediction models nor use a set of health issue-specific care quality measures for each health issue. Additionally, with time constraints, we may not be able to build a personalized search tool for IHPs that works well for all health issues. A good strategy for building the search tool is to focus on a few common health issues, particularly at the early stage of the tool development process.

Diabetes, heart disease (such as myocardial infarction and congestive heart failure), asthma, and chronic obstructive pulmonary disease (COPD) are four chronic health issues. They are common, can lead to costly acute exacerbations when not managed properly, and have low-cost treatments including prescription drugs and life style changes [8, page 254]. These four health issues offer excellent opportunities for the right IHPs to make major differences and hence are potentially good ones for us to focus on.

Besides these four health issues, several other health issues that may be good to focus on include breast cancer, major mental health disorder, chronic renal failure, arthritis (e.g., osteoarthritis) [42], lymphoma, inflammatory bowel disease, cystic fibrosis, HIV/AIDS, hypertension, alcohol/drug dependence, depression [40], hyperlipidemia, cervical cancer, colon cancer, and thyroid disease. Most of these other health issues have been studied before in the healthcare cost prediction literature.

We can use diagnosis codes to identify people with particular health issues [40]. For ICD-9-CM diagnosis codes, we can use the first three digits that indicate the diagnostic category [8, page 41]. Nevertheless, it is often insufficient to use one three-digit ICD-9-CM diagnosis code per health issue, as different manifestations of the same health issue could be classified across more than one three-digit ICD-9-CM diagnosis code. In this case, we

should use published medical literature for a comprehensive list of all three-digit ICD-9-CM diagnosis codes corresponding to the specific health issue [42].

4.12 Gender-specific and age-specific methods

Males are different from females. Children, adults under 65, and seniors are also different from each other. Each gender or age group has its own care quality measures and characteristics. For example, since children generally have low use of hospital care, hospital-based care quality measures are less useful for children than for adults [9, page 381]. As a second example, among the people with no health issues this year, the females of childbearing age, the oldest seniors, and the others all have rather different average healthcare costs next year [77].

For the same health issue, the outcomes and/or the important features for predicting a specific criterion's value may vary by the gender or age group [9, page 39]. For example, while diagnoses are often important risk factors for seniors, developmental milestones, functional status, family supports, and social environment may be more important risk factors for children [9, page 381].

As a second example, older patients often have different expectations of healthcare services and are more satisfied than younger patients [9, page 39, 19].

As a third example, as mentioned in Duncan *et al.* [8, page 25, 9, page 37], there is great variance in the healthcare cost for the same health issue at different ages. For certain health issues, such as diabetes, older patients have a higher average healthcare cost than younger patients because older patients often experience longer recuperative periods and more complications. The situation is reversed for some health issues, such as breast cancer, because older patients frequently choose to receive less aggressive treatment than younger patients.

Due to the differences among various age groups and genders, a single method for personalized searching for IHPs may not work well for all age groups and genders. In the personalized search tool for IHPs, it can be beneficial to use different methods, such as prediction models, for differing age groups and/or genders [77].

4.13 Grouper models

There are many health issues represented by the over 17,000 ICD-9-CM diagnosis codes, many medical procedures represented by the over 25,000 ICD-9-CM procedure codes, and many types of drugs represented by the over 317,000 NDC drug codes [58]. In predicting patient healthcare cost, care quality measures, or satisfaction, we would have an extremely large number of features if each health issue, medical procedure, and type of drug is used as a feature. Then the amount of available data would be insufficient for training the prediction models.

To reduce the number of features, researchers have developed various grouper models for health issues, medical procedures, and drugs, respectively [8, Chapter 5,

41, 42, 61]. These grouper models are described in detail in the risk adjustment literature. Several health insurance plans, e.g., Medicare, use grouper models to build healthcare cost prediction models for determining the plans' payments [42]. In predicting patient healthcare cost, care quality measures, or satisfaction, the personalized search tool for IHPs should use a grouper model for health issues, medical procedures, and drugs, respectively. In the rest of this section, we focus our discussion on the grouper models for health issues. The discussion on the grouper models for medical procedures and drugs [8, Chapter 5, 61] is similar and thus omitted.

A typical grouper model for health issues works in the following way [42]. All health issues are grouped into a moderate number of health issue categories based on clinical similarity and resource use. Health issue categories, rather than individual health issues, are used in building prediction models. The health issue categories are arranged into multiple hierarchies based on disease severity. Within a hierarchy, each health issue category represents a manifestation of the same health condition. If one category is an ascendant of another category, the manifestation represented by the ascendant category is more severe than that represented by the descendant category. For any patient whose health issues fall into more than one category, only those categories at the highest level of each hierarchy are used. For example, if a patient has two health issues falling into two different categories in the same hierarchy with one category a descendant of another, the ascendant category will be used while the descendant category will be omitted. This is to avoid repeatedly counting the same health condition multiple times.

Currently, several grouper models for health issues exist, including the Johns Hopkins Adjusted Clinical Groups (ACG) system, the Chronic Disease and Disability Payment System (CDPS), the Clinical Risk Groups (CRG) system, and the Diagnostic Cost Groups (DCG) system [42]. Some of these grouper models, e.g., the ACG system, are freely available, while the others require a fee to access. These grouper models have similar performance [42]. Hence, the personalized search tool for IHPs can choose to use any of them based on convenience, such as being freely available.

4.14 Exclusion of certain types of health issues and drugs

As mentioned in Ash *et al.* [77, 9, page 78, 51, 69], some health issues are medically insignificant, transitory, or definitively treated. These health issues should be excluded from the models for predicting patient healthcare cost, care quality measures, and satisfaction in the future. One such health issue is appendicitis that will never recur once the appendix is removed.

Similarly, as mentioned in Zhao *et al.* [63], certain types of drugs are known to have little or no predictive power based on clinical judgment and should be excluded from the prediction models. These drugs include those used for

major diagnostic testing, available in over-the-counter forms, or commonly used for various minor health issues.

4.15 Diagnosis codes

Sometimes, a diagnosis code is included on a medical claim because the health issue corresponding to the diagnosis code needs to be confirmed or ruled out, e.g., by ordering a diagnostic test [8, page 42, 9, page 125, 45]. Hence, if a diagnosis code for a health issue appears on a patient's medical claim, it does not necessarily mean that the patient has the health issue. According to diagnosis coding guidelines, this phenomenon usually occurs in the inpatient setting, but not in the outpatient setting [9, page 125].

To use a patient's medical claims data to identify the health issues of the patient, researchers have developed various clinical identification algorithms [8, Chapter 4]. Intuitively, the more times a diagnosis code appears on the patient's medical claims, the more likely the patient has the health issue corresponding to the diagnosis code. Each clinical identification algorithm is a set of rules constructed based on this intuition and attempts to strike a balance between sensitivity and specificity in identifying the patient's health issues.

In our application of matching patients with IHPs, it would be ideal if we can determine each patient's exact health issue(s). However, this is not mandatory. Even if we do not know each patient's exact health issue(s), we can use the patients' diagnosis codes in matching patients with IHPs. For example, we try to identify which IHPs are good at managing patients who either have or are suspected of having a specific health issue. If matching using diagnosis codes works well, the use of clinical identification algorithms will not be essential in the personalized search tool for IHPs [9, page 51].

The diagnosis codes on the same medical claim are not equally reliable. They can be given different priorities reflecting their reliability. As mentioned in Duncan *et al.* [8, pages 42 and 66, 9, page 103], on a medical claim, the first-listed (principle) diagnosis code is usually more reliable than any other following diagnosis codes because other diagnosis codes are often added to drive higher reimbursement. The diagnoses from the Evaluation and Management (E&M) Current Procedural Terminology (CPT) procedure codes are reliable because the E&M codes indicate that the IHP is treating the patient rather than conducting exploratory testing.

4.16 Features

Many patient features have been used in the past in predicting patient healthcare cost, care quality measures, and satisfaction. Each feature is either a raw attribute or a transformation of one or more raw attributes. As mentioned in Section 4.6, only some of these features are available to the personalized search tool for IHPs and hence can be used in the search tool's prediction models. In the following, we

provide a list of patient features that we have seen before in the research literature.

The features that are raw attributes are classified into the following six categories [9, page 31]:

- (1) Genetics:
 - genetically determined traits
 - genetic predisposition to specific health issues
 - genetic predisposition to specific health-related behaviors
 - family medical history
- (2) Demographic characteristics:
 - age
 - gender
 - race
 - ethnicity. Zip codes and surnames can be used to impute missing data on race and ethnicity [9, page 45].
 - primary language
 - level of English proficiency
 - country of origin
 - immigration status
- (3) Clinical factors:
 - acute physiological stability
 - diagnoses and other health issues, including present on admission (POA) indicators [9, page 125, 59]
 - age at diagnosis [78, page 181]
 - extent and severity of diagnoses
 - sensory functioning (vision, hearing)
 - physical functional status, such as difficulties in activities of daily living and instrumental activities of daily living [68]
 - breast feeding status
 - allergies
 - falls during the previous year [68]
 - cognitive status
 - mental and emotional health, such as presence of depression [68]
 - orders, such as medications, labs, exams, immunizations, imaging, counseling, and other procedures
 - vital signs including body temperature, systolic/diastolic blood pressure, pulse/heart rate, respiratory rate
 - other measures, exam results, and lab test results, such as spirometry reading, SpO₂, pain score, peak flow, and glucose level. Following the usual convention of handling missing data in test results, a normal value is imputed for each absent test result [64].
- (4) Psychosocial, socioeconomic, and environmental factors:
 - familial characteristics
 - household composition
 - education level
 - health literacy

- marital or partner status
- sexual orientation
- domestic violence
- economic resources and household income
- employment
- occupation
- housing
- neighborhood characteristics
- urban or rural residence
- geographic region
- health insurance plan/coverage, underinsurance
- disposition of patient, such as short-term hospital and skilled nursing facility [9, page 98]

For unavailable socioeconomic information, data on socioeconomic status of populations by zip code or US Census tract can be used as proxies [9, page 389].

(5) Health-related behaviors and activities:

- tobacco use, packs/day, number of years of smoking
- diet and nutrition
- obesity and overweight
- physical activity
- sleep
- excessive alcohol use
- illicit drug use
- unsafe sexual practices
- indicators for sedentary status, such as physical activity and walking for at least 30 minutes a week [68]
- other life style-related factors, such as stress, seat belt use, scuba diving, auto racing [8, page 6]

(6) Quality of life, attitudes, and perceptions:

- perception of overall health status and quality of life
- cultural beliefs and behaviors
- religious beliefs and behaviors
- preferences and expectations for healthcare services
- do not resuscitate, “comfort measures only” choices for end-of-life care

In general, we should avoid using any raw attribute that has either the floor effect or the ceiling effect [9, page 60, 79, page 135]. The floor effect refers to the case when data cannot take on a value lower than a specific number and thus, e.g., we cannot distinguish between the ill and the very ill. The ceiling effect is the opposite.

Many count measurements defined over a certain time period reflect a disease’s severity by capturing the intensity of healthcare resource usage (treatment intensity) [75]. The features that are transformations of one or more raw attributes include:

- the number of unique medications [61, 75]. As mentioned in Duncan [8, page 53], patients using multiple drugs (polypharmacy) are statistically more risky.
- the number of physician or hospital visits [75]

- the number of (chronic) health issues [43, 50, 75]
- the number of hospital claims [75]
- the number of procedures [61]
- the frequency of a procedure that involves multiple treatments [79, Chapter 10]
- the number of organ systems involved in the patient’s (chronic) diagnoses
- the number of difficulties in activities of daily living and instrumental activities of daily living [68]
- the number of health issues not treated before enrollment [68]
- indicator of emergency department visit [43]
- indicator of inpatient admission [43]
- combinations of health issues. The Centers for Medicare & Medicaid Services defines several health issue combinations that merit an additional risk score, such as diabetes and congestive heart failure [8, page 227].
- indicator of whether a patient’s healthcare costs exhibit a spike pattern [61]. As mentioned in Bertsimas *et al.* [61], the indicator is typically computed by checking the patient’s monthly healthcare costs to see whether the ratio of variance to mean is larger than a pre-determined threshold. The underlying intuition is that constant, relatively high healthcare costs represent a typical pattern for a patient with a chronic health issue and have a strong tendency to repeat in the future. In comparison, a patient whose past healthcare costs exhibit a spike pattern may have a low risk of incurring high healthcare costs in the future, as the spike usually results from an acute health issue that gets resolved in a short period of time.
- temporal patterns of health issues [56, 58, 61]
- temporal patterns of healthcare service usage [61]

Besides the features mentioned above, some other features that have not been used before in the research literature can also be useful for building the prediction models in the personalized search tool for IHPs. One such feature is how long a patient has been diagnosed with a chronic health issue. For instance, a newly diagnosed diabetic patient needs to be managed in a different way from a long-term diabetic patient with many complications. Another such feature is the frequency that an IHP’s patients switch to some other IHPs within the same clinic or hospital. The frequency partially and indirectly reflects how many patients are dissatisfied with the IHP. Several other such features are computed using the attributes in the IHP’s profile, as described in Luo [12].

4.17 Generalizing the idea of using IHP profile attributes to construct features

In building a personalized search tool for IHPs, one critical idea is to use IHP profile attributes to construct features and increase the prediction accuracy of patient care quality, healthcare cost, and satisfaction [12]. This idea is useful for multiple applications rather than only the search tool.

First, the improved predictions can enhance the performance of other healthcare applications relying on the predictions. One such healthcare application is to automatically identify candidate patients for case/care management programs [39, 51, 61, 68, 82]. Ideally, the candidate patients should be those with the poorest prognosis and/or expected to incur the highest healthcare costs in the future. For patients currently enrolled in a case/care management program, we periodically predict who will have good health outcomes and incur low healthcare costs in the future. These patients are likely not to need the case/care management program going forward. Another such healthcare application is to automatically identify patients who are likely to have poor health outcomes (e.g., be (re)admitted) in the future, and provide them with earlier follow-up appointments or nurse home visits to help prevent (re)admissions.

Second, in addition to the factors of patient care quality, healthcare cost, and satisfaction, the features constructed from IHP profile attributes can be used to improve the prediction accuracy of other factors such as missed appointments [87] and patient adherence [88, 89]. By identifying the patients at high risk for missing appointments, we can target resources, such as telephone and print reminders, to reduce the number of missed appointments [87]. By identifying the patients at high risk for non-adherence, we can take proactive actions to improve adherence [88]. Example actions include using longer-lasting medications to reduce the frequency of medication dosages, changing the route of medication delivery (e.g., from injection to oral), providing additional patient coaching, telephone reminders, and/or follow-up visits, and offering financial assistance such as copay discounts.

Third, this idea can be generalized and reused in multiple other applications for personalized care. The key observation underlying our idea is that a patient's care quality, healthcare cost, and degree of satisfaction depend not only on the patient's properties but also on the properties of the entity managing the patient. An IHP is one kind of such entity. Other kinds of such entity include a team of IHPs (e.g., a PCP and a dietitian), a type of IHP (e.g., dietitian), a healthcare facility/system, and a treatment model. Two example attributes of a healthcare facility are availability of specialty care (e.g., diabetes nurse educator, wound specialist) and resources (e.g., on-site lab, X-ray, case management). An example treatment model is the sequence and/or intensity of the treatments used. Another example treatment model is related to group therapy, in which multiple patients participate in an intervention program together. For group therapy, the treatment model includes the size and composition of the patient group as well as the duration of the intervention program.

By replacing IHPs with entities, we can expand our ability to match users with other entities. More specifically, for each entity, we can build a profile whose extended component aggregates the historical data of all of the

patients managed by the entity. We predict for each entity the user's care quality measures, healthcare cost, and degree of satisfaction if the user is going to be managed by the entity. We use entity profile attributes to construct features and increase the prediction accuracy of patient care quality, healthcare cost, and satisfaction. The features constructed for the other entities can be different from the features constructed for the IHPs. The predictions are used to identify an effective entity for the user.

5. Issues specifically related to predicting healthcare costs

In this section, we discuss several issues specifically related to predicting healthcare costs.

5.1 Performance metric

As mentioned in Ash and McCall [42], the performance of a healthcare cost prediction method can be evaluated using one of multiple metrics. No single performance metric is suitable for all healthcare applications and comprehensive in distinguishing the performance differences among various healthcare cost prediction methods. Instead, a proper performance metric should be chosen based on the nature of the specific healthcare application.

The most widely used performance metric for predicting healthcare costs is R^2 , which is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^n (a_i - \hat{a}_i)^2}{\sum_{i=1}^n (a_i - \bar{a})^2}.$$

Here, n is the number of people. a_i is the actual healthcare cost of the i -th person ($1 \leq i \leq n$). \hat{a}_i is the predicted healthcare cost of the i -th person. \bar{a} is the average actual healthcare cost of all n people.

Another commonly used performance metric is the Cumming's Prediction Measure (CPM), which is defined as

$$\text{CPM} = 1 - \frac{\sum_{i=1}^n |a_i - \hat{a}_i|}{\sum_{i=1}^n |a_i - \bar{a}|}.$$

CPM is essentially equivalent to the Mean Absolute Prediction Error (MAPE), which is defined as

$$\text{MAPE} = \frac{\sum_{i=1}^n |a_i - \hat{a}_i|}{n}.$$

R^2 is easily inflated by large prediction errors that often occur on high healthcare costs. Healthcare costs frequently have large outliers, the number and nature of which vary dramatically from one data set to another. Thus, R^2 tends to be unstable across different data sets. This can undesirably cause the optimal healthcare cost prediction method, which is selected from multiple candidate methods according to R^2 , to differ on different data sets. MAPE does not have these shortcomings. In our application of matching patients with IHPs, we need to use a performance metric, preferably a stable one, to select an optimal healthcare cost prediction method from multiple candidate methods. MAPE is likely to be better than R^2 for this selection purpose.

5.2 Prediction accuracy

As mentioned in Section 4.1, one or more criteria are used in ranking IHPs, i.e., in computing the degree of matching between an IHP's profile and the user's needs. When the user's healthcare cost is the only criterion, it may not matter much whether for each IHP we can predict precisely the user's healthcare cost if the user is going to be managed by the IHP [9, page 288]. Rather, what matters is the order of the predicted healthcare costs of the user for the IHPs. As long as the order of the predicted healthcare costs is correct, we can sort the IHPs in the right order. In contrast, the prediction accuracy of the user's healthcare cost matters more when IHPs are ranked based on the combination of the user's healthcare cost and other criteria.

Existing healthcare cost prediction methods usually predict the average healthcare cost for a group of patients satisfactorily [39, 45]. However, these prediction methods perform poorly on at the individual patient level, with an R^2 smaller than 20% [45] and an average prediction error roughly as large as the average healthcare cost [42]. This level of prediction accuracy is insufficient for our application of matching patients with IHPs.

Assuming that healthcare costs follow a linear statistical model, Newhouse *et al.* [80] estimated the maximum R^2 that can be achieved in predicting healthcare costs. The estimated value is relatively small. In reality, healthcare costs are far from following a linear statistical model [45]. Hence, the maximum accuracy that we can possibly achieve in predicting healthcare costs should be higher than the estimate provided in Newhouse *et al.* [80].

To improve the prediction accuracy of an individual patient's healthcare cost, we can proceed in multiple ways simultaneously. First, as mentioned in Frees *et al.* [73], we can use separate prediction models for different types of healthcare costs, e.g., inpatient healthcare cost and outpatient healthcare cost. Since outpatient visits are more frequent than inpatient admissions, outpatient healthcare costs are easier to predict than inpatient healthcare costs.

Second, some healthcare costs result from accidents and acute conditions, are difficult to predict, and hence greatly impact the average healthcare cost prediction accuracy on all of the patients [42, 51]. Nevertheless, over 75% of healthcare costs are due to chronic conditions [28]. Typically, a patient with one or more chronic conditions has recurring healthcare costs over a long period of time [40]. These healthcare costs tend to be predictable [42, 80]. Moreover, many patients with chronic conditions are sensitive patients because IHP effects will accumulate over time and have a substantial impact on their functioning in the long run [5].

The diagnosis codes of chronic conditions can be identified using some publicly available software [50, 81]. In the personalized search tool for IHPs, we can build a healthcare cost prediction model specifically for patients with chronic conditions, or alternatively, for patients with diagnosis codes of chronic conditions [42, 80]. On average, we would expect to achieve higher healthcare cost

prediction accuracy on these patients than on the other patients. This will help us find the right IHPs for these patients.

Third, as mentioned in Sections 4.7, 4.10, 4.12, and 4.16, we can use advanced, nonlinear machine learning methods [61], health issue-specific models, gender-specific models, age-specific models, and new features. In particular, several features that are computed using the attributes in the IHP's profile [12] have never been used before in any previous work.

Fourth, the healthcare cost of a patient with a specific health issue is the cost of treating all of his/her health issues [42]. To consider this factor, a machine learning model for predicting patient healthcare cost often uses many features, e.g., one feature per health issue. This causes the model to easily run out of enough training data [78, page 102].

To address this issue, we can use nonlinear, health issue-specific models, and reduce the number of features by grouping multiple less important features into one feature while keeping important features separate. More specifically, consider a specific health issue, such as asthma, and a patient with this health issue. We would expect the other health issues to fall into two classes. The first class includes a few health issues, such as the major ones that frequently co-occur with the specific health issue. Each health issue in the first class has a large impact on the patient's healthcare cost. The second class includes the many health issues that remain. Each health issue in the second class has a small impact on the patient's healthcare cost. Nevertheless, the machine learning model should not ignore these health issues, as they together have a large impact on the patient's healthcare cost.

We use medical knowledge and clinical judgment to identify these two classes of health issues. We group the health issues in the second class into several health issue categories based on clinical similarity and resource use, in a way similar to that used in traditional grouper models [42], but with further aggregation. Each health issue in the first class serves as a feature. For each health issue category, rather than each individual health issue, in the second class, we use one feature: the number of the patient's health issues in the category. This helps keep the total number of features moderate. We use these features to build nonlinear machine learning models for the specific health issue. Compared to the existing models using many features, these models use fewer features and are less likely to run into the insufficient training data problem.

5.3 Attributes related to the IHP's practice style

As explained in Ash and Byrne-Logan [40], a patient's previous healthcare cost is a powerful predictor of the patient's future healthcare cost, but is not used in some of the existing healthcare cost prediction models. This is due to the nature of the healthcare applications for which these prediction models are developed. The previous healthcare cost of a patient reflects both his health issues and the practice style of his IHP. In some healthcare applications, it

is undesirable to consider the IHP's practice style. Thus, no attribute related to the IHP's practice style is used. For example, health insurance plans use healthcare cost prediction models to determine their payments. They do not want to pay for the extra costs resulting from an IHP's suboptimal practice style. In other words, they are willing to pay more when the patient's health issues are generally known to require more care, rather than simply because more expensive care was provided to the patient in the past.

The case with our application of matching patients with IHPs is different. In our application, we would like to identify which IHPs' practice styles are effective for a particular patient. For each IHP, the previous healthcare costs of her patients reflect her practice style and thus should be used in building her profile [12]. This information can be useful for predicting a patient's healthcare cost, care quality measures, and/or degree of satisfaction with the IHP if the patient is going to be managed by the IHP. For example, due to the increasing burden of healthcare costs, patients pay more attention to healthcare costs these days than before. Consequently, patients may tend to be more satisfied with the IHPs prescribing less costly care [4]. As another example, many high healthcare costs result from unnecessary care, which may lead to worse health outcomes, e.g., because of iatrogenic harm via overtreatment [43].

Besides the previous healthcare cost of a patient, several other attributes are also related to his IHP's practice style and hence not used in some of the existing healthcare cost prediction models. Examples of these attributes include the number of times that the patient is hospitalized and the frequency that a specific diagnosis appears on the patient's medical claims [77]. In general, all of these attributes can be used in the personalized search tool for IHPs.

5.4 Top-coding

Usually, extremely high healthcare costs occur on only a small portion of patients. Many existing healthcare cost prediction methods use the top-coding technique to limit these patients' potentially significant influence on the predicted values [9, page 264, 42]. In top-coding, each healthcare cost higher than a pre-determined constant value C (e.g., \$25,000) is replaced by C . Top-coding is most useful for those healthcare applications that do not require predicting as precisely as possible the healthcare cost of each patient who will have a high healthcare cost.

One such healthcare application is to identify patients who are likely to incur high healthcare costs and then enroll them in case management programs [39, 51, 61, 68, 82]. The case management programs can help defer or avoid adverse outcomes and subsequently reduce patient healthcare costs. The application cares whether a particular patient is likely to incur high healthcare cost, not exactly how much his healthcare cost is going to be.

Another such healthcare application is to predict the average healthcare cost of a group of patients, which is

essential for a health insurance plan in determining its payments. As mentioned in Iezzoni [9, page 265], in this application top-coding lowers the average healthcare cost by only a very small fraction. The lost dollars can be put back into the analysis by multiplying each predicted value by a factor: the mean of the original healthcare costs divided by the mean of the top-coded healthcare costs.

In the personalized search tool for IHPs, the predicted healthcare cost is combined with other criteria to compute the degree of matching between an IHP's profile and the user's needs [12]. To better compute the degree of matching, it is desirable to predict as precisely as possible the healthcare cost of each patient who will have a high healthcare cost. Top-coding eliminates the differences among differing high healthcare costs, creates difficulty for making such a prediction, and hence is likely to be useless for the search tool.

For example, consider two IHPs with different practice styles and three patients who are almost identical in every aspect. The first patient is managed by the first IHP with a healthcare cost of $2C$. The second patient is managed by the second IHP with a healthcare cost of $3C$. We would expect the healthcare cost of the third patient to be $2C$ and $3C$ if he is going to be managed by the first IHP and the second IHP, respectively. Thus, everything else being equal, the first IHP is likely to manage the third patient more effectively than the second IHP. Nevertheless, if top-coding with a threshold of C is used, both $2C$ and $3C$ will be replaced by the same value C . Then we cannot determine which IHP is likely to manage the third patient more effectively.

5.5 Claims run-out

As mentioned in Duncan [8, page 43], a lag of several months typically exists between the date of service and adjudication of a medical claim. If data is extracted too soon after service, some healthcare cost data will be missing because not all of the medical claims have been paid yet. In predicting healthcare costs, people often wait for a certain number of days, such as 90 days, to obtain a relatively complete medical claims data set.

As mentioned in Section 5.4, in the healthcare application of identifying patients who are likely to incur high healthcare costs and then enrolling them in case management programs, we do not have to predict each patient's healthcare cost as precisely as possible. Therefore, waiting for most or all of the medical claims to get paid is unnecessary [8, page 46].

In our application of matching patients with IHPs, each patient's healthcare cost data will be affected in roughly the same way if we do not wait for most or all of the medical claims to get paid. This is likely to have minimum impact on the order in which the IHPs are sorted, particularly for the few top-ranked IHPs. Thus, we would expect that waiting for most or all of the medical claims to be paid is still unnecessary, although this time the underlying reason

is different from that in the healthcare application of identifying patients for a case management program.

5.6 Billed, allowed, or reimbursed cost

Each medical claim is associated with a billed cost, an allowed cost, and a reimbursed cost [8, page 43]. For predicting healthcare costs, it is better to use allowed cost because allowed cost is less subject to variation due to member cost-sharing [8, page 45].

5.7 Adjustments for annual cost increase and regional price differences

Due to inflation and use of more expensive and advanced technology, healthcare costs increase annually even if everything else remains the same each year [8, page 19, 9, page 443]. This annual cost increase needs to be considered in building a healthcare cost prediction model. To adjust for this annual cost increase and make healthcare costs from two consecutive years comparable with each other, a typical method used is to multiply each healthcare cost in year 1 by a factor: the mean of the healthcare costs in year 2 divided by the mean of the healthcare costs in year 1 [8, page 19].

Similarly, regional price differences need to be considered in building a healthcare cost prediction model. To make healthcare costs from different regions comparable with each other, we can adopt one of the following two methods. The first method is to use geographic adjustment factors to reflect price differences across the country [83]. The second method is to use some proxies for healthcare costs, such as relative value units (RVUs) [83, 84, 85], that measure consumption of healthcare services and are unaffected by regional price differences [9, page 115, 45]. The second method works for Veterans Health Administration (VHA) data, which does not contain dollar claims for healthcare services because the VHA uses global budgeting [9, page 115]. Since HMOs receive capitated payments from Medicare and do not submit individual claims to Medicare [22], the second method is also needed in analyzing the healthcare costs of seniors who are enrolled in both HMO and Medicare.

6. Issues specifically related to predicting care quality measures

In this section, we discuss several issues specifically related to predicting care quality measures.

6.1 Outcome measures vs. process measures

There are three types of care quality measures: outcome measures, process measures, and structure measures [9, page 28, 86]. Outcome measures evaluate how patients fare, such as a patient's blood pressure level, whether a patient is readmitted within a certain time period, and whether a patient becomes dead within a certain time period. Process measures look at what is done for patients, e.g., whether a myocardial infarction patient receives an aspirin

prescription on discharge. Structure measures assess physical equipment and facilities.

Both outcome measures and process measures can help differentiate the performance of the IHPs in the same clinic and are readily available, e.g., from the provider's electronic medical record system. Structure measures do not have these properties. The personalized search tool for IHPs uses both outcome measures and process measures.

Many process measures do not directly measure the effectiveness and appropriateness of care [48]. For example, a process measure may give an IHP credit for providing advice on smoking cessation, regardless of whether the advice is careless. Consequently, outcome measures are more important than process measures [86]. In computing the degree of matching between an IHP's profile and the user's needs, outcome measures should be given higher priorities or weights than process measures.

As mentioned in Berenson *et al.* [48], intermediary outcome measures should be used with caution. For example, treating patients intensively to lower their hemoglobin A1C levels may not help them achieve desired final outcomes.

6.2 Generic care quality measures vs. health issue-specific care quality measures

Each health issue has its own set of care quality measures. Some care quality measures are generic ones, while the others are specific to the health issue. As mentioned in Kane and Radosevich [79, Chapter 7], a generic care quality measure may miss some clinically significant treatment effects and not cover some necessary dimensions of health. Health issue-specific care quality measures can help make up these issues and complement generic care quality measures. For instance, some generic care quality measures are designed to distinguish the health status of reasonably healthy people from that of sick people. They cannot distinguish between the ill and the very ill.

Lists of health issue-specific care quality measures are provided in several dedicated books [79, page 152] as well as scattered in many journal articles. The personalized search tool for IHPs should use both generic care quality measures and health issue-specific care quality measures. Similarly, health issue-specific prediction models should use both generic features and health issue-specific features.

6.3 Mental health outcome measures

Hundreds of outcome measures exist for mental health, but there is little consensus on which outcome measures should be used for which patients [9, page 394]. This creates some difficulty for using the personalized search tool for IHPs on mental health patients. In practice, we do our best and use whatever mental health outcome measure is available in the data set.

7. Issues specifically related to predicting patient satisfaction

In this section, we discuss several issues specifically related to predicting patient satisfaction.

7.1 Effect of the care setting

Patient satisfaction is one criterion used in computing the degree of matching between an IHP's profile and the user's needs. The care setting affects the criterion's effectiveness at differentiating the performance of different IHPs and hence the criterion's priority. As mentioned in Jackson *et al.* [19], patients seen in a continuity care setting usually report being fully satisfied more than 90% of the time. In this case, it is difficult to use patient satisfaction to effectively differentiate the performance of different IHPs. In contrast, there is greater variation in patient satisfaction in a walk-in care setting, where patients are randomly assigned to IHPs based on IHP availability. In that case, patient satisfaction is a more effective criterion for differentiating the performance of different IHPs.

8. Conclusions

Personalized search for IHPs is a new topic. This paper surveys various issues on designing a personalized search tool for IHPs that are not covered in our previous paper. We outline some preliminary thoughts on how to address these issues. To fully address these issues, much more research work is needed. We hope this paper can stimulate future research work on this new topic.

A personalized search tool for IHPs is useful for both consumers and healthcare providers. Consumers can use the search tool to find satisfactory IHPs. Likewise, if an IHP realizes that he cannot manage a specific patient well or is asked to make a referral, he can use the search tool to refer the patient to another IHP who is likely to be good at managing the patient. Moreover, through the data mining process used in the search tool, we can better understand the reason why an IHP treats a specific patient successfully or unsuccessfully. For instance, we can use process measures to construct features. Then we identify both treatment patterns leading to good outcomes and treatment patterns leading to poor outcomes. The information contained in the treatment patterns can provide hints on how to improve every IHP's performance and ultimately patients' outcomes [90].

Acknowledgments

We thank Leslie A. Lenert, Selena Thomas, Libin Shen, Lewis J. Frey, Zac E. Imel, Farrant Sakaguchi, Susan Terry, Katherine Sward, Peter J. Haug, Kensaku Kawamoto, Bryan L. Stone, Bruce E. Bray, Qing T. Zeng, and Maureen Murtaugh for helpful discussions.

References

1. RAND Health. Consumers and health care quality information: need, availability, utility. <http://www.chcf.org/~media/MEDIA%20LIBRARY%20Files/PDF/C/PDF%20ConsumersAndHealthCareQualityInformation.pdf>, 2001.

2. Stewart, M.A., Effective physician-patient communication and health outcomes: a review. *CMAJ* 1995; 152(9): 1423-1433.
3. Safran, D.G., Taira, D.A., and Rogers, W.H. *et al.*, Linking primary care performance to outcomes of care. *J Fam Pract.* 1998; 47(3): 213-220.
4. Westgate, A., Patient satisfaction and care quality: Problems and predictions. Available at <http://www.physicianspractice.com/blog/patient-satisfaction-and-care-quality-problems-and-predictions>, 2012.
5. Baldwin, S.A., and Imel, Z.E., Therapist effects: Findings and methods. In: Lambert MJ, editor. *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*, 6th ed. Hoboken, NJ: Wiley; 2013.
6. Luborsky, L., McLellan, A.T., and Woody, G.E. *et al.*, Therapist success and its determinants. *Arch Gen Psychiatry* 1985; 42(6): 602-11.
7. McKay, K.M., Imel, Z.E., and Wampold, B.E., Psychiatrist effects in the psychopharmacological treatment of depression. *J Affect Disord.* 2006; 92(2-3): 287-90.
8. Duncan, I., *Healthcare Risk Adjustment and Predictive Modeling*. Winsted, CT: ACTEX Publications Inc.; 2011.
9. Iezzoni, L.I., *Risk adjustment for measuring health care outcomes*, 4th ed. Chicago, IL: Health Administration Press; 2013.
10. Laska, K.M., Smith, T.L., and Wislocki, A.P. *et al.*, Uniformity of evidence-based treatments in practice? Therapist effects in the delivery of cognitive processing therapy for PTSD. *J Couns Psychol.* 2013; 60(1): 31-41.
11. The IHI Triple Aim. <http://www.ihl.org/offerings/Initiatives/TripleAim/Pages/default.aspx>, 2013.
12. Luo, G., Open issues in intelligent personal health record - an updated status report for 2012. *J Med Syst.* 2013; 37(3):9943.
13. Teevan, J., Dumais, S.T., and Liebling, D.J., To personalize or not to personalize: modeling queries with variation in user intent. *Proceedings of SIGIR'08*, pp. 163-170, 2008.
14. Melucci, M., Contextual search: A computational framework. *Foundations and Trends in Information Retrieval* 2012; 6(4-5): 257-405.
15. Han, J., Kamber, M., and Pei, J., *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA: Morgan Kaufmann; 2011.
16. Ricci, F., Rokach, L., and Shapira, B. *et al.*, *Recommender Systems Handbook*. New York, NY: Springer; 2010.
17. Li, J., Fu, A.W., and He, H. *et al.*, Mining risk patterns in medical data. *KDD* 2005: 770-775.
18. Pacific Business Group on Health. Advancing physician performance measurement: Using administrative data to assess physician quality and efficiency. Available at http://www.pbgh.org/storage/documents/reports/PBGHP3Report_09-01-05final.pdf, 2005.
19. Jackson, J.L., Chamberlin, J., and Kroenke, K., Predictors of patient satisfaction. *Soc Sci Med.* 2001; 52(4): 609-20.
20. Saxon, D., and Barkham, M., Patterns of therapist variability: therapist effects and the contribution of patient severity and risk. *J Consult Clin Psychol.* 2012; 80(4): 535-46.
21. Brown, G.S., Lambert, M.J., and Jones, E.R. *et al.*, Identifying highly effective psychotherapists in a managed care environment. *Am J Manag Care* 2005; 11(8): 513-20.
22. The Dartmouth Atlas of Health Care. <http://www.dartmouthatlas.org/data/topic/topic.aspx?cat=21>, 2013.

23. Gifford, E., and Foster, E.M., Provider-level effects on psychiatric inpatient length of stay for youth with mental health and substance abuse disorders. *Med Care* 2008; 46(3): 240-6.
24. Kramer, T.L., Daniels, A.S., and Ziemann, G.L. *et al.*, Psychiatric practice variations in the diagnosis and treatment of major depression. *Psychiatr Serv.* 2000; 51(3): 336-40.
25. Elstein, A.S., Shulman, L.S., and Sprafka, S.A., Medical problem solving: A ten-year retrospective. *Evaluation & the Health Professions* 1990; 13(1): 5-36.
26. Elstein, A.S., Shulman, L.S., and Sprafka, S.A., *Medical Problem Solving: An Analysis of Clinical Reasoning.* Cambridge, MA: Harvard University Press; 1978.
27. Milstein, A., and Lee, T.H., Comparing physicians on efficiency. *N Engl J Med.* 2007; 357(26): 2649-52.
28. Chronic diseases - the power to prevent, the call to control: at a glance 2009. Available at <http://www.cdc.gov/chronicdisease/resources/publications/aag/chronic.htm>, 2009.
29. Imel, Z.E., Baldwin, S., and Atkins, D.C. *et al.*, Racial/ethnic disparities in therapist effectiveness: a conceptualization and initial study of cultural competence. *J Couns Psychol.* 2011; 58(3): 290-8.
30. Adams, J.L., Mehrotra, A., and Thomas, J.W. *et al.*, Physician cost profiling - reliability and risk of misclassification. *N Engl J Med.* 2010; 362(11): 1014-21.
31. Medicare Physician Compare homepage. <http://www.medicare.gov/physiciancompare/search.html>, 2013.
32. Medicare Hospital Compare homepage. <http://www.medicare.gov/hospitalcompare/search.html>, 2013.
33. Medicare Nursing Home Compare homepage. <http://www.medicare.gov/nursinghomecompare/search.html>, 2013.
34. Medicare Home Health Compare homepage. <http://www.medicare.gov/homehealthcompare/search.html>, 2013.
35. Sitzia, J., and Wood, N., Patient satisfaction: a review of issues and concepts. *Soc Sci Med.* 1997; 45(12): 1829-43.
36. Wampold, B.E., and Brown, G.S., Estimating variability in outcomes attributable to therapists: a naturalistic study of outcomes in managed care. *J Consult Clin Psychol.* 2005; 73(5): 914-23.
37. Baldwin, S.A., Wampold, B.E., and Imel, Z.E., Untangling the alliance-outcome correlation: exploring the relative importance of therapist and patient variability in the alliance. *J Consult Clin Psychol.* 2007; 75(6): 842-52.
38. Imel, Z.E., Hubbard, R.A., and Rutter, C.M. *et al.*, Patient-rated alliance as a measure of therapist performance in two clinical settings. *J Consult Clin Psychol.* 2013; 81(1): 154-65.
39. Moturu, S., Johnson, W., and Liu, H., Predicting future high-cost patients: A real-world risk modeling application. *Proceedings of BIBM'07*, pp 202-208, 2007.
40. Ash, A.S., and Byrne-Logan, S., How well do models work? Predicting health care costs. *Proceedings of the Section on Statistics in Epidemiology, American Statistical Association* 1998, 42-49.
41. Moturu, S.T., Johnson, W.G., and Liu, H., Predictive risk modelling for forecasting high-cost patients: a real-world application using Medicaid data. *Int. J. of Biomedical Engineering and Technology* 2010; 3(1/2): 114-132.
42. Ash, A., and McCall, N., Risk assessment of military populations to predict health care cost and utilization. Available at http://www.rti.org/pubs/tricare_riskassessment_final_report_combined.pdf, 2005.
43. Fenton, J.J., Jerant, A.F., and Bertakis, K.D. *et al.*, The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. *Arch Intern Med.* 2012; 172(5): 405-11.
44. Hussey, P.S., de Vries, H., and Romley, J. *et al.*, A systematic review of health care efficiency measures. *Health Serv Res.* 2009; 44(3): 784-805.
45. Diehr, P., Yanez, D., and Ash, A. *et al.*, Methods for analyzing health care utilization and costs. *Annu Rev Public Health* 1999; 20: 125-44.
46. Balasubramanian, H., Banerjee, R., and Denton, B. *et al.*, Improving clinical access and continuity through physician panel redesign. *J Gen Intern Med.* 2010; 25(10): 1109-15.
47. Diwakar, G., and Brian, D.T., Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions* 2008; 40(9): 800-819.
48. Berenson, R.A., Pronovost, P.J., and Krumholz, H.M., Achieving the potential of health care performance measures. Timely analysis of immediate health policy issues. Urban Institute report available at <http://www.rwjf.org/content/dam/farm/reports/reports/2013/rwjf406195>, 2013.
49. Hornbrook, M.C., Hurtado, A.V., and Johnson, R.E., Health care episodes: definition, measurement and use. *Med Care Rev.* 1985; 42(2): 163-218.
50. Fleishman, J.A., and Cohen, J.W., Using information on clinical conditions to predict high-cost patients. *Health Serv Res.* 2010; 45(2): 532-52.
51. Fleishman, J.A., Cohen, J.W., and Manning, W.G. *et al.*, Using the SF-12 health status measure to improve predictions of medical expenditures. *Med Care* 2006; 44(5 Suppl): I54-63.
52. Stone, E.M., Heinold, J.W., and Ewing, L.M. *et al.*, Accessing physician information on the Internet. http://www.commonwealthfund.org/~media/Files/Publications/Fund%20Report/2002/Jan/Accessing%20Physician%20Information%20on%20the%20Internet/stone_mdinternet_503%20pdf.pdf, 2002.
53. Health Maintenance Organization (HMO) Plan. <http://www.medicare.gov/sign-up-change-plans/medicare-health-plans/medicare-advantage-plans/hmo-plans.html>, 2013.
54. The APCD (all-payer claims database) Council homepage. <http://www.apcdouncil.org/>, 2013.
55. ResDAC homepage. <http://www.resdac.org/>, 2013.
56. Davis, D.A., Chawla, N.V., and Christakis, N.A. *et al.*, Time to CARE: a collaborative engine for practical disease prediction. *Data Min. Knowl. Discov.* 2010; 20(3): 388-415.
57. Mancuso, D., Nordlund, D.J., and Felver, B.E., Using claims-based risk indicators to predict health outcomes for children with mental illness. RDA Report 3.38. Available at <http://www.dshs.wa.gov/pdf/ms/rda/research/3/38.pdf>, 2013.
58. Shenk, K.N., Markuzon, N., and Bertsimas, D.J., Patterns of heart attacks. Master's thesis, Massachusetts Institute of Technology, 2010. Available at <http://dspace.mit.edu/handle/1721.1/61198>.

59. Pine, M., Jordan, H.S., and Elixhauser, A. *et al.*, Enhancement of claims data to improve risk adjustment of hospital mortality. *JAMA* 2007; 297(1): 71-6.
60. Perkins, A.J., Kroenke, K., and Unützer, J. *et al.*, Common comorbidity scales were similar in their ability to predict health care costs and mortality. *J Clin Epidemiol.* 2004; 57(10): 1040-8.
61. Bertsimas, D., Bjarnadóttir, M.V., and Kane, M.A. *et al.*, Algorithmic prediction of health-care costs. *Operations Research* 2008; 56(6): 1382-1392.
62. Kansagara, D., Englander, H., and Salanitro, A. *et al.*, Risk prediction models for hospital readmission: a systematic review. *JAMA* 2011; 306(15): 1688-98.
63. Zhao, Y., Ash, A.S., and Ellis, R.P. *et al.*, Predicting pharmacy costs and other medical costs using diagnoses and drug claims. *Med Care* 2005; 43(1): 34-43.
64. Escobar, G.J., Greene, J.D., and Scheirer, P. *et al.*, Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. *Med Care* 2008; 46(3): 232-9.
65. Powers, C.A., Meyer, C.M., and Roebuck, M.C. *et al.*, Predictive modeling of total healthcare costs using pharmacy claims data: a comparison of alternative econometric cost modeling techniques. *Med Care* 2005; 43(11): 1065-72.
66. Sun, Y., Heng, B.H., and Tay, S.Y. *et al.*, Predicting hospital admissions at emergency department triage using routine administrative data. *Acad Emerg Med.* 2011; 18(8): 844-50.
67. Wikehult, B., Ekselius, L., and Gerdin, B. *et al.*, Prediction of patient satisfaction with care one year after burn. *Burns* 2009; 35(2): 194-200.
68. Anderson, R.T., Balkrishnan, R., and Camacho, F., Risk classification of Medicare HMO enrollee cost levels using a decision-tree approach. *Am J Manag Care.* 2004; 10(2 Pt 1): 89-98.
69. Pope, G.C., Kautter, J., and Ellis, R.P. *et al.*, Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financ Rev.* 2004; 25(4): 119-41.
70. Robinson, J.W., Regression tree boosting to adjust health care cost predictions for diagnostic mix. *Health Serv Res.* 2008; 43(2): 755-772.
71. Yoo, I., Alafaireet, P., and Marinov, M. *et al.*, Data mining in healthcare and biomedicine: a survey of the literature. *J Med Syst.* 2012; 36(4): 2431-48.
72. Kuhn, M., and Johnson, K., *Applied Predictive Modeling.* New York, NY: Springer; 2013.
73. Frees, E.W., Gao, J., and Rosenberg, M.A., Predicting the frequency and amount of health care expenditures. *North American Actuarial Journal* 2011; 15(3): 377-392.
74. Zhou, Z., *Ensemble Methods: Foundations and Algorithms.* Boca Raton, FL: CRC Press; 2012.
75. Farley, J.F., Harley, C.R., and Devine, J.W., A comparison of comorbidity measurements to predict healthcare expenditures. *Am J Manag Care* 2006; 12(2): 110-9.
76. Tabak, Y.P., Sun, X., and Derby, K.G. *et al.*, Development and validation of a disease-specific risk adjustment system using automated clinical data. *Health Serv Res.* 2010; 45(6 Pt 1): 1815-35.
77. Ash, A.S., Ellis, R.P., and Pope, G.C. *et al.*, Using diagnoses to describe populations and predict costs. *Health Care Financ Rev.* 2000; 21(3): 7-28.
78. Steyerberg, E.W., *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.* New York, NY: Springer; 2009.
79. Kane, R.L., and Radosevich, D.M., *Conducting Health Outcomes Research.* Sudbury, MA: Jones & Bartlett Learning; 2010.
80. Newhouse, J.P., Manning, W.G., and Keeler, E.B. *et al.*, Adjusting capitation rates using objective health measures and prior utilization. *Health Care Financing Review* 1989; 10: 41-54.
81. Gorina, Y., and Kramarow, E.A., Identifying chronic conditions in Medicare claims data: evaluating the Chronic Condition Data Warehouse algorithm. *Health Serv Res.* 2011; 46(5): 1610-27.
82. Meenan, R.T., Goodman, M.J., and Fishman, P.A. *et al.*, Using risk-adjustment models to identify high-cost risks. *Med Care* 2003; 41(11): 1301-12.
83. Johnson, S.E., and Newton, W.P., Resource-based relative value units: a primer for academic family physicians. *Fam Med.* 2002; 34(3): 172-6.
84. Hsiao, W.C., Braun, P., and Dunn, D. *et al.*, Results and policy implications of the resource-based relative-value study. *N Engl J Med.* 1988; 319(13): 881-8.
85. Hsiao, W.C., Braun, P., and Yntema, D. *et al.*, Estimating physicians' work for a resource-based relative-value scale. *N Engl J Med.* 1988; 319(13): 835-41.
86. Duke University Medical Center. Measurement: Process and outcome indicators. http://patientsafetyed.duhs.duke.edu/module_a/measurement/measurement.html, 2013.
87. Chariatte, V., Berchtold, A., and Akre, C. *et al.*, Missed appointments in an outpatient clinic for adolescents, an approach to predict the risk of missing. *J Adolesc Health.* 2008; 43(1): 38-45.
88. Saha, Y., and Singal, R., Predicting patient adherence: why and how. <http://www.cognizant.com/InsightsWhitepapers/Predicting-Patient-Adherence-Why-and-How.pdf>, 2013.
89. Passanante, M., Understanding and maximizing prescriber impact on patient compliance and persistency. <http://www.eyeforpharma.com/pcusa07/presentations/02MikePassanante.pdf>, 2007.
90. James, B.C., and Savitz, L.A., How Intermountain trimmed health care costs through robust quality improvement efforts. *Health Aff* 30(6): 1185-91, 2011.