

# Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework

Nitesh V. Chawla, PhD<sup>1</sup> and Darcy A. Davis, PhD<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, Interdisciplinary Center for Network Science and Applications (iCeNSA), University of Notre Dame, Notre Dame, IN, USA; <sup>2</sup>University of Notre Dame, Notre Dame, IN, USA.

Faced with unsustainable costs and enormous amounts of under-utilized data, health care needs more efficient practices, research, and tools to harness the full benefits of personal health and healthcare-related data. Imagine visiting your physician's office with a list of concerns and questions. What if you could walk out the office with a personalized assessment of your health? What if you could have personalized disease management and wellness plan? These are the goals and vision of the work discussed in this paper. The timing is right for such a research direction—given the changes in health care, reimbursement, reform, meaningful use of electronic health care data, and patient-centered outcome mandate. We present the foundations of work that takes a Big Data driven approach towards personalized healthcare, and demonstrate its applicability to patient-centered outcomes, meaningful use, and reducing re-admission rates.

**KEY WORDS:** personalized healthcare; data mining; patient-centered outcomes.

J Gen Intern Med 28(Suppl 3):S660-5

DOI: 10.1007/s11606-013-2455-8

© Society of General Internal Medicine 2013

“All of life and human relations have become so incomprehensibly complex that, when you think about it, it becomes terrifying and your heart stands still.”<sup>1</sup>

If Russian author and physician Anton Chekhov was overwhelmed by the complexity of life in 1897, imagine how he might feel about medicine in modern times. He practiced medicine in the early years of the progressive era of medicine, when diseases were still just phenotypes and preventive concepts like nutrition, exercise, hygiene, and public sanitation were radical new ideas.<sup>2</sup> Today, the phenomenology of disease includes a massive systematized knowledge of molecular interactions, influenced both by genetics and environmental factors. Furthermore, the biomedical structures underlying disease overlap and have complex effects on one another. This complex system or network view of disease can be overwhelming, but also serves as a rich resource for understanding and improving human life. Researchers in the emerging fields of bioinformatics and computational medicine, in particular, seek

analytic tools to bring order and understanding among the complexity to keep hearts pumping, rather than making them stand still. The potential for such applications to enhance medical decision-making and healthcare delivery are no less visible in the developing world than in developed economies. As information systems for health become globally available, as they are today in the IU-Kenya AMPATH program, massive repositories of personal health information will also emerge.<sup>3</sup> Data mining might assist to address various questions as they arise in global health setting, such as: what is the likelihood that this patient will develop resistance to her antiretroviral regimen? Experience ART side effects? Develop Kaposi's sarcoma?

Despite the massive influx of data created by rapid advances in genomic technologies and increasing collection of clinical data, we have a very incomplete picture of disease at the “systems” level.<sup>4</sup> These rich sources of data have potential for an increased understanding of disease mechanisms and better healthcare, but the size, complexity, and biases of the data present many challenges. There is a recognizable need for scalable computational tools that can discover patterns without discounting the statistical complexity of heterogeneous data or falling prey to the noise it includes.

*Data-driven and networks-driven thinking and methods can play a critical role in the emergence of personalized healthcare.* Numerous diseases have preventable risk factors or at least indicators of risk. Elucidating these disease characteristics may help in personalized healthcare, and help reduce disease burden. However, the possible combination of risk factors is so complex, it's impossible for an individual physician to fully analyze it (in real time) at the time of patient interaction. Currently, providers take careful histories and do physical examinations and selective laboratory testing to determine patient health and risk for future disease. These are generally limited to a few diseases and by the skill and knowledge of the individual provider and competing priorities for individual visits. Thus, to take next big steps in personalized healthcare requires a computing and analytics framework to aggregate and integrate big data, discover deep knowledge about patient similarities and connections, and provide personalized disease risk profiles for each individual patient, derived from not only the electronic medical record information of

Published online June 25, 2013

that patient, but also from similarities of that patient to millions of other patients. This opens the opportunities for proactive medicine, actively managing disease, empowering the patients, and effectively leading to a reduction in re-admission rates. This is the thesis our work on personalized healthcare.

Currently, the focus on personalized healthcare is based on genomic revolution. Advances in technology have provided an extensive list of mutations, SNPs and the subsequent likelihood of developing specific diseases.<sup>5</sup> The presumption is that once all disease-related mutations are catalogued, we will be able to predict individual susceptibility based on various molecular biomarkers. However, our ability to identify mutations has greatly outstripped our understanding of their clinical relevance.<sup>6,7</sup> Moreover, many of these mutations and disease associated SNPs give relatively weak signals; not all individuals with specific genetic risks develop disease. Application of genome-based approaches is likely to take considerable time for the relationships to be understood. Fortunately, there are other approaches that may allow application of personalized health care that do not depend on waiting mature understanding and use of genomic approaches.

We hypothesize that phenotype and disease history-based approaches offer the promise of rapid advances towards personalized disease prediction, management, and wellness. In a corollary development, we note that disease data are also particularly well suited to a network representation; the structure emerges naturally since biological phenotypes are products of molecular interaction.<sup>8</sup> Thus, genetic and molecular data can be integrated with phenotypic data to improve disease modeling and understanding, when available.<sup>9</sup> Leveraging clinically reported traits and symptoms, along with the biological information of diseases and their interactions, can provide a summary of possible risk factors, underlying causes, and anticipated comorbidities.<sup>9</sup>

In this work, we present an overview of our prior and existing work<sup>9-12</sup> on the role of Big Data analytics and computation in healthcare, and the potential it holds for transformations in personalized healthcare and biomedical discovery.

## PERSONALIZED HEALTHCARE: A DATA-DRIVEN APPROACH

Healthcare is moving from a disease-centered model towards a patient-centered model.<sup>13</sup> In a disease-centered model, physicians' decision making is centered around the clinical expertise, and data from medical evidence and various tests. In a patient-centered model, patients actively participate in their own care and receive services focused on individual needs and preferences, informed by advice and oversight from their healthcare providers. At the same time

as this patient-centered care model is being emphasized in health care delivery, the potential for 'personalizing' health care from a disease prevention, disease management, and therapeutics perspective is increasing. Healthcare informatics and advanced analytics (or data science) may contribute to this shift from population-based evidence for health care decision-making to the fusion of population and individual-based evidence in health care.

Our work is motivated by patient-centric model that creates a personalized disease risk profile, as well as a disease management plan and wellness plan for an individual.<sup>10-12</sup> We situate our work on the observation that diseases do not occur in isolation. They result from an interaction between genetic, molecular, environmental, and lifestyle factors.<sup>5,6,14</sup> Similarities in lifestyles and experiences matter, along with genetic predispositions, in our risk for diseases. Patients exposed to similar risk, lifestyle and environmental factors may have similar outcomes. How to deeply leverage the "big data" resident in electronic medical records, patients' experiences and histories, to create a personalized disease risk profile for an individual patient? Can we develop a patient-centered model for personalized care to answer questions such as: What diseases am I at risk for developing? How should I manage them? What wellness strategies may best work for me?

We approached this problem by learning from the work on 'collaborative filtering methodology' used in other settings by recommendation systems.<sup>15-18</sup> Essentially, collaborative filtering is a data mining technique designed to predict a user's opinion about an item or service based on the known preferences of a large group of users.<sup>19</sup> Most applications assume that the input is partial preference information for each user. That is, the user's opinion or rating is known for a few items, but usually unknown for a strong majority of the item set. The basic principle behind collaborative filtering is that users who have similar taste on some items are likely to have similar taste on other items. This information is then used to make personalized predictions on the movies one may want to watch (such as in Netflix.com) or books one may want to read (such as in Amazon.com)

To that end, we posited that the problem of patient-centered and personalized disease risk profile is analogous to recommendation systems used for movies or books. We are trying to leverage similarities across a large group of patient pool, in real-time, to deliver a personalized plan to an individual. It is known that similarity in lifestyle and environmental factors, and genetic predispositions cause us to be susceptible to similar diseases.<sup>20</sup> Co-occurring factors have a synergistic effect, leading to unexpectedly high risk.<sup>6</sup> Using collaborative filtering, we can generate predictions on other diseases based on a set of other similar patients. However, there are a number of challenges. The diseases do not have a rating system or a "preferential scoring" as in movies or books. We know if a person had a disease in the past or does not have a disease (rather not been diagnosed

with the disease); there is no preferential ranking or rating provided by individuals. The challenge of absent disease cannot be ignored as that could simply imply that the patient has not been, yet, diagnosed with the disease.

Thus, we are able to incorporate a vast array of disease comorbidities, which are effectively leveraged for personalized disease prediction, management plan, and wellness for an individual patient. Our work leverages the similarities and shared experiences among individuals in a healthcare system and beyond (potentially millions of individuals) on: patient history, disease timing, disease progression, prognosis, and wellness strategies. This big data is filtered to result in the personalized plan. We now introduce our framework called the Collaborative Assessment and Recommendation Engine (CARE) for patient-centered disease prediction and management.<sup>10,11</sup>

### CARE: A PATIENT-CENTRIC PERSONALIZED HEALTHCARE FRAMEWORK

Imagine visiting your physician’s office with a list of concerns and questions. What if you could walk out the office with a personalized assessment of your health, along with a list of personalized and important lifestyle change recommendations based on your predicted health risks? What if the physician was afforded a limitless experience to gauge the impact of your diseases towards developing other diseases in the future? What if you could find out that there are other patients similar to you not only with respect to major (more common symptoms), but also with respect to rare issues that have puzzled your respective doctor? What if you could have others’ experiences at your fingertips and fathom the lifestyle changes warranted for mitigating the disease(s)? These are the goals of our work in developing the patient-centric and personalized disease risk prediction model using collaborative filtering techniques. We have developed a system called Collaborative Assessment and Recommendation Engine (CARE) for personalized disease risk predictions.<sup>10–12</sup> At the core of CARE is a novel collaborative filtering method that captures patient similarities and produces personalized disease risk profiles for individuals.

A number of computer-aided methods have been developed for disease prediction. For example, APACHE III is a prognostic scoring system for predicting inpatient mortality.<sup>21</sup> There are also other disease specific models for specific conditions, such as heart conditions,<sup>22</sup> digestive disorders,<sup>23</sup> hepatitis,<sup>24</sup> Alzheimer’s disease,<sup>25</sup> and cancer.<sup>26</sup> Our approach is distinctly different from existing work in that we are trying to build a general predictive system that can utilize a less constrained feature space, i.e. taking into account all available demographics and previous medical history. Our work addresses the criteria of prospective

healthcare espoused by Snyderman et al.,<sup>27</sup> where they call for the creation and validation of new models for determining disease risk and suggest that data mining is a “central feature” of prospective healthcare. Using Big Data science, specifically collaborative filtering, we generate predictions focused on other diseases that are based on data from similar patients. These predictions can lead to better management and prevention strategies, and potentially empower the patient to have a dialogue that leads to an improved wellbeing. They may also provide guidance for relatively rare diseases and complications that could elude a physician, but are elucidated by the data-driven integration of experiences of many physicians and patients.

Figure 1 illustrates the theoretic platform for CARE.<sup>11</sup> When an individual arrives in an office with his or her medical history, this medical history is compared with all the other patients’ medical histories that one may have access to based on defined similarity constraints. The similarity constraints could be defined on the basis of shared diseases, symptoms, family histories, lab results, urban/rural residencies, occupation, demographics, etc. Based on the similarity computation, a pool of patients most similar to the patient under consideration (also called the active patient) is selected. Once the similar universe of patients is selected, we apply collaborative filtering using inverse frequency and vector similarity (Breese et al.,1998). The functioning of collaborative filtering can be specified mathematically. Traditionally, collaborative filtering is used to make a prediction  $p(a,j)$  for user  $a$ , the active user, for an item  $j$ , based on the similarity between user  $a$  and every other user  $i$ , who has previously rated that item or expressed a preference

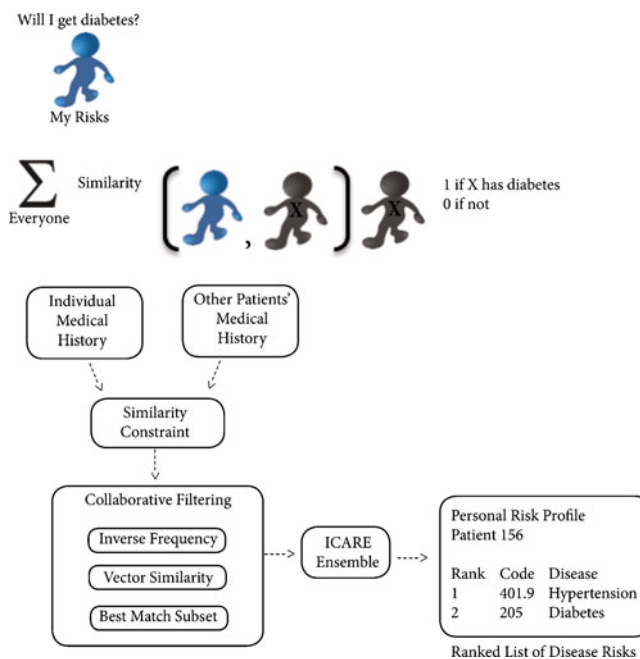


Figure 1. CARE process diagram

for that item ( $v_{i,j}$ ) for that item. The entire training set of items is defined as  $I$ , and  $I_j$  is the subset of users that have rated the item  $j$ . The similarity  $w(a,i)$  between users  $a$  and  $I$  is calculated by vector similarity; that is,

$$w(a,i) = \frac{v_{a,j} v_{i,j}}{\sqrt{\sum_{k \in I_a} v_{a,k}^2} \sqrt{\sum_{k \in I_i} v_{i,k}^2}}$$

$I_j$  is the set of items rated by user  $i$ . The prediction score takes into account the average vote  $\bar{v}_i$  of each user to account for personal differences. The normalizing constant  $\alpha$  is added so that the sum of weights is equal to 1, constraining the prediction within the range of possible votes. Thus, the general collaborative filtering equation is:

$$p(a,j) = \bar{v}_a + \alpha \sum_{i \in I_j} w(a,i) (\bar{v}_i - \bar{v}_i)$$

However, this equation will not work in the medical domain. The user in this case is a patient and the items are diseases. Each patient  $i$  either has the disease  $j$  or does not have the disease  $j$ . There is no preference or rating. It is thus a binary problem (1 or 0). Therefore, we have to incorporate binary ratings and also remove the effect of the range of ratings, expressed in the computation of  $\bar{v}_i$ . The modified general equation instead incorporates the random expectation of each disease given a population sample, referred to as  $\bar{v}_j$ . Specifically,  $\bar{v}_j = |I_j|/|I|$ . Thus, the  $p(a,j)$  can now be expressed in the CARE framework as,

$$p(a,j) = \bar{v}_j + \alpha \sum_{i \in I_j} w(a,i) (1 - \bar{v}_j)$$

$p(a,j)$  is the probability that an active patient ( $a$ ) will develop a disease ( $j$ ) in the future, where an active patient is the testing patient (or the patient on whom the predictions are being applied) and disease  $j$  is the disease that the patient has not been diagnosed with to date. Intuitively, the equation treats the random expectation  $\bar{v}_j$  as the baseline expectation of each patient having disease  $j$  and adds additional risk based on the similarity to other patients with disease  $j$ . Now, we found that the more common diseases dominated the similarity computation, as expected, since, the patients having one or more common diseases share a bigger overlap. To that end, we incorporated the inverse frequency metric that gives higher weight to the fact that two patients share a relatively rare disease. There can be many medical diagnoses shared among patients, but most important contributions may arise uncommon connections. The inverse frequency of disease  $j$  is defined as  $f_i = \log n n_n$ , where  $n$  is the number of patients in the training set and  $n_j$  is the number of patients that have disease  $j$ . This sets up the basis of collaborative filtering. We found that inverse frequency was insufficient for

countering the domination of common diseases and suppression of rare conditions. Even though rare matches were scored highly, there were many more common weak matches. To counter this, we built an ensemble of collaborative filtering models for each active patient, which we call the iterative version of CARE (ICARE).<sup>11</sup> Essentially, we take one disease at a time, find the group of patients that have that disease in common, apply collaborative filtering in that group, and generate  $p(a,j)$ . We then have a set of  $p(a,j)$ , expressed as the risk of patient  $a$  developing disease  $j$  in the future. Each of the  $p(a,j)$  is contributed to by different disease clusters and collaborative filtering method that collectively make an ensemble. We then take the maximal  $p(a,j)$  for specifying the probability of disease  $j$  on patient  $a$ . Thus, ICARE is essentially an ensemble of CARE models, as it first identifies cluster of patients based on a single disease patient similarity, then runs CARE on that particular cluster, and repeats the process for each unique disease in the active patient's history. Since each of these iterations is completely independent of each other, we can easily process the iterations simultaneously on a distributed or parallel system. We note that while each disease may result in a different cluster (patient group), we still used all the entire past disease history of the active patient  $a$ .

To demonstrate the concept, consider the following case. A patient has disease  $x$  and disease  $y$ , and CARE is making a prediction on disease  $j$ . Disease  $x$  is 100 times more prevalent than disease  $y$ . Disease  $j$  co-occurs in 2 % of the patients with disease  $x$  and in 50 % of the patients with disease  $y$ . Using the CARE clustering method, disease  $j$  should be 2.5 % prevalent in the cluster. Any disease that co-occurs with 3 % or more of patients with disease  $x$  will have an advantage. Realistically, this patient should have *at least* a 50 % risk of developing disease  $j$ . By using ICARE, essentially an ensemble of CARE models, this strong link can be isolated despite the relative rarity of disease  $y$ . In the first iteration (member of the ensemble), the cluster consists only of patients with disease  $x$ , and disease  $j$  has a cluster baseline of 0.02. This will not lead to a strong prediction,  $p(a,j)$ . However, in the second iteration (second member of the ensemble), the cluster consists of patients with disease  $y$ , giving disease  $j$  a cluster baseline of 0.5. This results in a much fairer risk assessment. Although ICARE clusters diseases individually, vector similarity with inverse frequency is still performed using the full history vector to develop predictability on diseases that connect to multiple past diagnoses. For example, consider a disease  $k$  expressed by 50 % of patients with disease  $x$  and none with disease  $y$ . Diseases  $j$  and  $k$  will both have a maximum cluster baseline of 0.5, but disease  $j$  will have a higher prediction score since it links to both disease  $x$  and  $y$ .

In order to reduce the number of predictions and the runtime of the ensembles, we only predict on diseases for

which the cluster baseline is significantly higher than the population baseline ( $\overline{v_j}$ ). That is, if the population baseline is similar to the cluster baseline, then the disease being predicted on does not have a good set of predictive diseases in the cluster. We refer the reader to our prior work for complete technical details on CARE and ICARE.<sup>11</sup> We will refer to the overall system as CARE in our discussion.

*The result is a ranked list of diseases ranked in order from highest to lowest risk. These lists can be simplified into a shorter, less specific version by grouping diagnoses.* In our published work,<sup>10–12</sup> we validated CARE on a Medicare database of 13 million patients with 32 million visits, spread over 4 years.<sup>28</sup> Each data record represents a hospital visit, represented by a patient ID and a list of up to ten diagnosis codes, as defined by the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). The International Statistical Classification of Diseases and Related Health Problems (ICD) provide codes to classify disease and a wide variety of signs, symptoms, abnormal findings, social circumstances, and external causes of injury or disease.

We deployed a leave-one-patient-out, leave-one-visit-out validation strategy. That is, consider an active patient  $a$  on whom CARE will generate predictions about possible disease risks in the future. The patient  $a$  has visited the physician once, and has its first visit diagnoses as the medical history (disease vector) for input to CARE. This disease vector of  $a$  is used for similarity computation against all the patients in the database, resulting in  $p(a,j)$  for each disease  $j$  that is not in the  $a$ 's medical history. The diseases are ranked based on their  $p(a,j)$ . These predictions are then checked for correctness in the future visit(s) of the patient  $a$ . That is, did the patient  $a$  develop the said disease by visit 2, 3, etc. We computed how many of the diseases that a patient  $a$  develops in the future are ranked in the top 20 or top 100. This effectively allows us to evaluate the precision of CARE in the top- $k$  (where  $k$  can be 10 or 20 or 100). The goal is to provide a physician with a short list that includes high-risk diseases of a patient. In addition to the coverage, we also provided the average rank of diseases. ICARE was able to capture about 51 % of future diseases (coverage) in the top 20 with an average rank of 5.755. This is a list of manageable size that provides early warning indicators of more than 50 % of diseases that a person may develop in the future. This is an encouraging result by just using ICD-9-CM codes. We conjecture that the result can significantly improve if additional information such as family history, genetic data, lab results, symptoms can also be included, and this is part of our ongoing work.

## SUMMARY

The CARE system was developed to serve as a data-driven computational aid for physicians assessing the disease risks

facing their patients. The compelling future disease coverage of CARE represents early warning indicators of potential disease risks of an individual, which can then be converted in to a dialogue between the physician and patient, leading to patient empowerment. In its most conservative use, the rank lists can provide reminders for conditions that busy doctors may have overlooked. Utilized to full potential, CARE can be used to explore broader disease histories, suggest previously unconsidered concerns, and facilitate discussion about early testing and prevention, as well as wellness strategies that may ring a more familiar bell with an individual and are essentially doable. *We believe that our work can lead to reduced re-admission rates, improved quality of care ratings, can demonstrate meaningful use, impact personal and population health, and push forward the discussion and impact on the patient-centered paradigm.* We are expanding our work to include additional data such as labs, symptoms, etc., beyond the ICD-9-CM codes (that can be limiting). However, utilization of ICD codes by CARE allows for a seamless integration with a variety of electronic health care systems that use or will embrace the standard of such diagnoses codes. Finally, we are setting up test-and-control samples for evaluating CARE and incorporating feedback. We are working with physician and patient focus groups to establish gold standard tests of CARE disease predictions, as well as usability of the system. *With the increase in the use of electronic medical records, CARE becomes an increasingly important possibility, bringing Big Data and data science to proactive and personalized patient care.*

---

**Conflict of Interest:** The authors declare that they do not have a conflict of interest.

**Corresponding Author:** Nitesh V. Chawla, PhD; Department of Computer Science and Engineering, Interdisciplinary Center for Network Science and Applications (iCeNSA), University of Notre Dame, Notre Dame, IN, USA (e-mail: nchawla@nd.edu).

## REFERENCES

1. **Chekhov AP, Matlaw RE.** Anton Chekhov's Short Stories. WW Norton & Company Incorporated; 2008.
2. **Engs RC.** The Progressive Era's Health Reform Movement: A Historical Dictionary. Praeger Publishers; 2003.
3. **Ampath.** <http://www.ampathkenya.org/>. Accessed 2012 Dec 20.
4. **Kitano H.** Systems biology: a brief overview. *Science*. 2002;295(5560):1662–1664.
5. **Burton PR, Clayton DG, Cardon LR, et al.** Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–678.
6. **Loscalzo J.** Association studies in an era of too much information. *Circulation*. 2007;116(17):1866–1870.
7. **Loscalzo J, Kohane I, Barabasi AL.** Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Mol Syst Biol*. 2007;3(1).

8. **Schadt EE.** Molecular networks as sensors and drivers of common human diseases. *Nature*. 2009;461(7261):218–223.
9. **Davis DA, Chawla NV.** Exploring and exploiting disease interactions from multi-relational gene and phenotype networks. *PLoS One*. 2011;6(7):e22670.
10. **Davis DA, Chawla NV, et al.** Predicting individual disease risk based on medical history. In *Proceedings of the 17th ACM conference on Information and knowledge management*. 2008:769–778. ACM.
11. **Davis DA, Chawla NV, Christakis NA, Barabási AL.** Time to CARE: a collaborative engine for practical disease prediction. *Data Min Knowl Disc*. 2010;20(3):388–415.
12. **Dentino B, Davis D, Chawla NV.** HealthCareND: leveraging EHR and CARE for prospective healthcare. In *Proceedings of the 1st ACM International Health Informatics Symposium*. 2010:841–844. ACM.
13. **Stanton MW.** Expanding patient-centered care to empower patients and assist providers. *Research in Action*. Agency for Healthcare Research and Quality, Rockville; 2002. Available: <http://www.ahrq.gov/qual/ptcareria.htm>.
14. **Barabási AL.** Network medicine—from obesity to the “diseasome”. *N Engl J Med*. 2007;357(4):404–407.
15. **Breese JS, Heckerman D, Kadie C.** Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*. 1998: 43–52. Morgan Kaufmann Publishers Inc.
16. **Goldberg K, Roeder T, Gupta D, et al.** A constant time collaborative filtering algorithm. *Inf Retr*. 2001;4(2):133–151.
17. **Herlocker JL, Konstan JA, Terveen LG, Riedl JT.** Evaluating collaborative filtering recommender systems. *ACM Trans Inf Syst (TOIS)*. 2004;22(1):5–53.
18. **Hofmann T.** Latent semantic models for collaborative filtering. *ACM Trans Inf Syst (TOIS)*. 2004;22(1):89–115.
19. **Su X, Khoshgoftaar TM.** A survey of collaborative filtering techniques. *Adv Artif Intell*. 2009;4:421–425.
20. **Starfield B, Lemke KW, Bernhardt T, et al.** Comorbidity: implications for the importance of primary care in ‘case’management. *Ann Fam Med*. 2003;1(1):8–14.
21. **Wong DT, Knaus WA.** Predicting outcome in critical care: the current status of the APACHE prognostic scoring system. *Can J Anesth Can J Anesth*. 1991;38(3):374–383.
22. **Jenkins KJ, Gauvreau K, Newburger K, et al.** Consensus-based method for risk adjustment for surgery for congenital heart disease. *J Thorac Cardiovasc Surg*. 2002;123(1):110–118.
23. **Clamp SE, Myren J, Bouchier IA, Watkinson G, de Dombal FT.** Diagnosis of inflammatory bowel disease: an international multicentre scoring system. *Br Med J (Clin Res Ed)*. 1982;284(6309):91.
24. **Piscaglia F, Cucchetti A, Benlloch S, et al.** Prediction of significant fibrosis in hepatitis C virus infected liver transplant recipients by artificial neural network analysis of clinical factors. *Eur J Gastroenterol Hepatol*. 2006;18(12):1255–1261.
25. **Liu YH, Chen SM, Lin CY, et al.** Motion tracking on elbow tissue from ultrasonic image sequence for patients with lateral epicondylitis. In *29th Annual International Conference of the IEEE*. 2007:95–98.
26. **Mould RF.** Prediction of long-term survival rates of cancer patients. *Lancet*. 2003;361(9353):262.
27. **Snyderman R, Langheier J.** Prospective health care: the second transformation of medicine. *Genome Biol*. 2006;7(2):104.
28. **Christakis NA, Allison PD.** Mortality after the hospitalization of a spouse. *N Engl J Med*. 2006;354(7):719–730.