Evidence Based Medicine Methodology Through Big Data Analysis

Ashish Kaushik
Dept. of Information Science and Technology
School of Engineering and Technology
Bengaluru, India

Sachin Kumar
Dept. of Computer Science and Technology
School of Engineering and Technology
Bengaluru, India

Manjunath C R
Dept. of Computer Science and Technology
School of Engineering and Technology
Bengaluru, India

Abstract:- Health care analysis poses a consistent challenge to physicians and is the area of research in which trillions of amounts are being spent by all countries. The patient health conditions are analyzed and exact disease is diagnosed using Big Data analytics with Evidence Based methodology. Patient analysis report is generated to monitor the recovery and patient feedback of the suggested drug is recorded to witness the success-rate of the diagnosis process. Proposing new idea for the patients on their medicine dosage by applying strictly age limit and providing nearby shops and Medicare Centre by their GPS location on their device. Here a probabilistic data collection mechanism is designed and the correlation analysis of those collected data is performed. A stochastic prediction model is designed to foresee the future health condition of the most correlated patients based on their current health status. A cloud-enabled big data analytic platform is the best way to analyze the structured and unstructured data generated from healthcare management systems. Patient's medical history and current evidences are considered to diagnose and drug suggestion. The EBM manifesto offered here grew from that awareness. It is an open invitation for others to contribute to and join a movement towards better evidence by providing a roadmap for how to achieve the listed priorities and to share the lessons from achievements already made.

Keywords:- Data analytics; Medicare; GP; EBM.

I. INTRODUCTION

Enormous information in medicinal services is an explanatory domain to deal with the gigantic volume of organized and unstructured patient information. As indicated by the investigators, the social insurance information volume has come to 150 Exabyte's in 2011 and has expanded to zettabyte scale in the present time. The wellbeing information are credited as large information, which is characterized by 5Vs as far as Volume, Velocity, Variety, Value, and Veracity. The gathered patient information are of peta or zeta bytes, which depict the volume. The speed is communicated things of information landing rate from the patients. Assortment clarifies the differentiated informational collections as for the organized, semi-organized and unstructured informational indexes, for example, clinical reports, EHRs, and radiological pictures and veracity clarifies the honesty of the informational collections regarding information accessibility and realness. The gathered information are changed into significant experiences, which clarify the esteem in 5Vs Evidence-based solution is the administration of individual patients through individual clinical aptitude incorporated with the devoted and prudent utilization of receding and flow best confirmation from clinical care inquire about. This approach considers missing, inadequate, or low-quality proof and requires the use of clinical judgement. The logical writing is the significant wellspring of confirmation for prove based prescription, in spite of the fact that writing based proof ought to regularly be supplemented by nearby, practice based proof for person also, site-particular clinical basic leadership.

Evidence based medicine (EBM) is a generally acknowledged worldview for restorative practice that includes the unequivocal utilization of ebb and flow best proof, that is, amazing patient-focused clinical research, for example, reports from randomized controlled trials, in settling on choices about patient care. Physiological information of patients are the essential and indispensable elements in human services enormous information investigative. Subsequently, legitimate crude information must be gathered with an effective way in a medicinal domain. A distributed computing empowered circulated stockpiling and handling condition is basic to store and process the human services information, which can be gotten to anyplace and whenever. Huge information scientific is connected in social insurance to recognize the bunches of patients, infections and future expectations with the assistance of different machine learning apparatuses. In a learning social insurance framework, information are examined and utilized as experiences constantly for understanding consideration. Amid this procedure, the patient information is joined with the clinical reports for machine learning tools. In a learning healthcare system, data are analyzed and used as insights continuously for patient care. During this process, the patient data are combined with the clinical reports for better recommendations and choices. So far restricted investigations have achieved among the patients taking different numbers of health parameters of same or different departments.

II. PROBLEM STATEMENT

Exact analysis of disease is a testing task and is habitually put off as a result of numerous components muddling malady finding. Huge life reserve funds can be accomplished if an exact analysis can be made. More dependable and capable Clinical decision support systems (CDSS) are required to decrease the time required for finding and increment the analysis exactness. These days the social insurance industry is confronting the issue on strategy for quality administrations at moderate expenses. In many rural areas due to lack of proper medical equipments and medical facilities patients remain deprived of actual treatment.

Insufficient knowledge about symptoms and medicine, patients end up consuming wrong medicine and resulting to side effect. Absence of replication since numerous essential investigations are not rehashed, or even 'repeatable', by other autonomous specialists to confirm the outcomes. Some significant investigations that have been duplicated have discovered shockingly extraordinary and even opposing outcomes.

The reward frameworks inside medical research, especially in the scholarly community, boosts amount of productions over nature of research. Patient's perception on medicines will be varied on their requirement so, clinicians have to be very expertise and able to understand situations. At emergency patient's fail to get medicine due to improper facilities like data about medicines, navigation for accurate clinic. Generally, in rural areas government fails to provide such facilities on time and patients have to suffer a lot. About 85% of spending on research currently goes wasted. HealthCare information systems easily support single organizational units but face trouble in exchange of information between the care units, the community and the care providers and at big scale. Electronic health record is stored in various different autonomous IT related repositories at local, and regional, and national level resulting in difficulty in finding relevant medical knowledge. EHR have been developed for storing patient information and are not integrated with general medical knowledge. Work activities are not facilitated thus consume from 50 to 80 percentage of the physician's time because its performed manually instead of computerized. "ease of use" factor is absent in current systems and making difficult to navigate preventing practitioners from achieving effective, efficient and satisfied goals. No easily compilation and communicate information to accounting and management systems at execution and principal levels. There is no patient engagement facility. They themselves cannot easily access care information Found after studies that providing patients more access to their wellbeing data can urge them to partake in their own particular care and can likewise self-deal with their wellbeing condition, find out about their medicinal issues, and enhance patient– supplier.

III. EXISTING SYSTEM

It is observed that the patients are suffering from the disease in huge numbers and one of the reason is due to lack of knowledge and details about the disease and its proper medicines. It happens due to lack of equipments in hospitals also and main reason is improper transport facilities. They have to go to the clinics, hospitals by themselves whether it's far or near, they have to manage and in rural areas people will not aware of their problems only they avoid, ignore and at last have to face death. It's very sad that our government and system fails in such scenarios, and not providing proper facilities. This is recommended that outcomes from half of all trials are never published according to a historic point audit, and that positive trials are twice as liable to be published as results from negative trials. Patients have to pay fee for even consultancy also and they do their time waste for travelling and searching for a good and fascinated center. 85% of research spending as of now goes to waste One third (34%) of researchers report flawed research works on, including data mining for factually noteworthy impacts, specific announcing

of results, exchanging results, production inclination, convention deviations, and hiding irreconcilable circumstances.

IV. LIMITATION OF THE EXISTING SYSTEM

The randomized clinical trial had been introduced as an intense instrument for evaluating the practicality and prosperity of medicines. The Evidence-Based Medicine improvement fought that restorative practice was exorbitantly subjective and searched for, making it difficult to propel more important reliance on appropriated look into. In doing, all things considered, it prescribed that clinical judgment and foolish reasoning are less trustworthy sorts of confirmation in the arrangement. In spite of the fact that this view seems to have been relinquished by the EBM development, it keeps on applying impact, bringing about disarray and discussion. The advancement of clinical trial outlines and strategies were gotten from the claim to fame of the study of disease transmission and this may, to a limited extent, clarify why the proof that they give was given need by the EBM development, which was to a great extent made out of individuals from that strength. Be that as it may, in spite of their qualities, the confinements of clinical trials are progressively perceived in applying proof in the down to earth universe of clinical medication. Despite the fact that the restrictions of clinical trials are not the blame of Evidence Based Medicine, an over-dependence on them, to the detriment of clinical proof. Limitations of the clinical trials and systematic reviews are due to:

- Unnatural of trial patients in terms of therapy, age and comorbidity.
- Over-reliance on statistical as opposed to clinical significance.
- Misleading comes about because of announcing inclination, unseemly unite of little trials, impact of changes in illness mortality, and forecast after some time.
- Unrepresentativeness of trial subjects.
- Statistical, as opposed to clinical significance.

V. PROPOSED SYSTEM

Our proposed system is aimed at collecting medicine information from hospitals and medical centers and applying big data analysis to the raw medical records and propose proper details of drugs, their quantity to be taken and timing for consuming. By enabling GPS technology patients will get the location of nearby medical stores where prescribed medicine is available. Enhancing the systematic use of existing evidence for best analysis and report results. Support innovation, quality change and security through the better utilization of genuine information. Providing the prescribed and exact list of medicines and their guidelines to patient's for better awareness and it's consumption hence,it reduce the overdosage problems and minimize the count of ill patients in hospitals. Patient Health Condition is broke down utilizing Bio Medical Investigations like Blood Sugar, Lipid profile, Height, Weight, BMI, BP, Heart Problems and different components are considered. These assembled information is part into little parts and handled in parallel to recognize the appropriate treatment and medications utilizing Hadoop Distributed File

System (HDFS). Map diminish idea of BigData is used to process the voluminous clinical information proficiently. Machine Learning strategies are connected for diagnosing the disease. We are aiming on an application too which gives this medicine information online through big data analysis and have feature like gps system and from map they will easily get the nearest Medicare Centre where they will get prescribed medicine based on their query and dosage details.

A. The Ebm Process

Generally, evidence based medicine consists of 7steps (Table 1) and as described below:-

Step 1. Identify Patient Condition:

The initial phase in Evidence-Based-Medicine is to satisfactorily analyze patient's clinical issue. A case of patient's condition might be - 55 years of age, 220 pounds weight and white man, with hypertension (BP), which is recently determined to have Type-2 diabetes. Other significant data incorporates: hypersensitivity condition; foot, kidney and eye wellbeing; cardio-vascular maladies and so forth. The patient's health condition can be comprehended by the EBM at various times finding, lab, and managerial information.

Step 2. Formulation of EBM Question:

After the condition of patient is recognized, this progression continues with integrating those condition stored into the clinical questions. A case of a clinical question might be: Is wholesome treatment as emotional as Oral Medication to the 55 years of age with overweight and T2 Diabetes patients, and also high Blood Pressure? At that point, clinical inquiries are changed over into EBM question. One of the approaches is through the Population-Intervention-Comparison-Outcome (PICO) criteria to structure the EBM question. PICO is getting noteworthy consideration as a way to look at the monetary, social and clinical result of various intercession options. Populace alludes to the statistic and clinical data of patient; intervention alludes to the conceivable course of action; while correlation alludes to looking at between elective intercessions, or between "intervention "and "no mediation"; at last, result implies yield which we need to get to (clinical, financial or social).

A case of Evidence-Based-Medicine question might be:
Population — Age:55; Weight: 220pounds; BP:
Intervention — Nutritional Therapy
Comparison — Oral Medication
Outcome — Control on Glucose

When all is said in done, the medicinal inquiry available can mention to determination, guess, treatment, iatrogenic harms, nature of care, or wellbeing financial matters.

Step 3. Gathering of Evidences:

This step focus around retrieving the data identified with the Evidence-Based-Medicine questions. This progression additionally incorporates preparatory assessment of significant data (individual article, or singular snippet of data from the electronic wellbeing records) before it is utilized for facilitate analysis. For an instance, if an article does not fulfill the base criteria, at that point it ought to be disposed of. The data is broken down, integrated, and converted into

consumable module that fathom the Evidence-Based-Medicine question. The wellsprings of confirmation could be electronic wellbeing records (Practice Based Evidences) as well as research studies (Literature Based Evidence).

Step 4. Evaluation of Evidences:

The training rules that are produced in past advance are further curated in this progression. Medicinal field is data basic area—remedy data help to settle on amend choice and spare life; in any case, broken data may welcome passing or confounds the patient's condition. Therefore, clinicians are hesitant to believe a PC program that show confirm without clear data about dependability and provenance of the data. Specialist and professionals should sufficiently assess the proof and keep up trust in prove. Proof assessment should be possible in three stages: review the individual articles, or individual snippet of data; review the general confirmation (in the wake of collecting the individual snippet of data); and ascertain the factual importance of general data. The yield of confirmation assessment is the review (A-D or comparative) and factual essentialness (p value) of proof.

Step 5. Conversion of the Critically Appraised Evidence into Consumable Unit (practicable to use at the point of care):

Here, the focus is on changing over fundamentally evaluated prove into consumable rules ideally in machine comprehensible arrangement, e.g. Evidence coded into the UMLS.

Step 6. Evidence Presentation and Use:

According to the evidence presentation, literature focuses on four key factors exhibit the EBM at the purpose of care; display no less than 2 options course of actions; show unwavering quality and reference alongside the evidence, fit data in a single screen of a PC.

Step 7. Evaluation of the result of evidence to put into practice:

This progression surveys if partners are really getting advantage by honing Evidence Based Medicine. Not just this, it likewise knows which prove is working, and what are the changes required in the knowledgebase of Evidence Based Medicine. At the point when innovation is utilized to convey Evidence Based Medicine, this progression ends up less demanding via automating the evaluation task.

A. The Ebm Process Table

Steps	Example of Evidence-Based-Medicine Related Tasks
1. Identify Patient Condition	Identify the complete Patient's condition by analyzing the past and the present health records which include symptoms, demographics, diagnosis, allergies, tests, economic background, post conditions for hereditary and also other related information.
2. Formulation of EBM Question	Synthesizing the clinical questions into Evidence-Based-Medicine question. PICO criteria (Population -Comparison -Outcome) is the one of the most common ways of generating the EBM question.
3. Generation of Evidences and Analysis.	Generation of evidences by integrating the information from the research studies(Clinical, RCT, literature, Guidelines),and practice dependent evidence.
4. Evaluation of Evidences	Grade individual pieces of information(individual article, or singular information from the practice based evidence).
5. Conversion of Evidences into Consumable Unit.	Converting critically approved evidence into easily accessible & usable format.
6. Evidence Presentation and Use.	Present minimum 2 alternatives actions course at the point of care.
7. Evaluate of Implementing EBM into practice.	Tracking of the success and failure of individual piece of information.

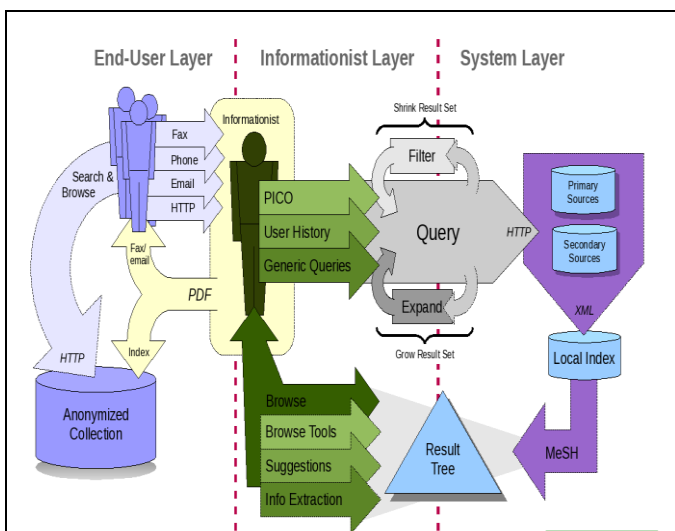*B. Architecture Diagram*

Fig 1:- Architecture of Proposed Evidence based medicine System
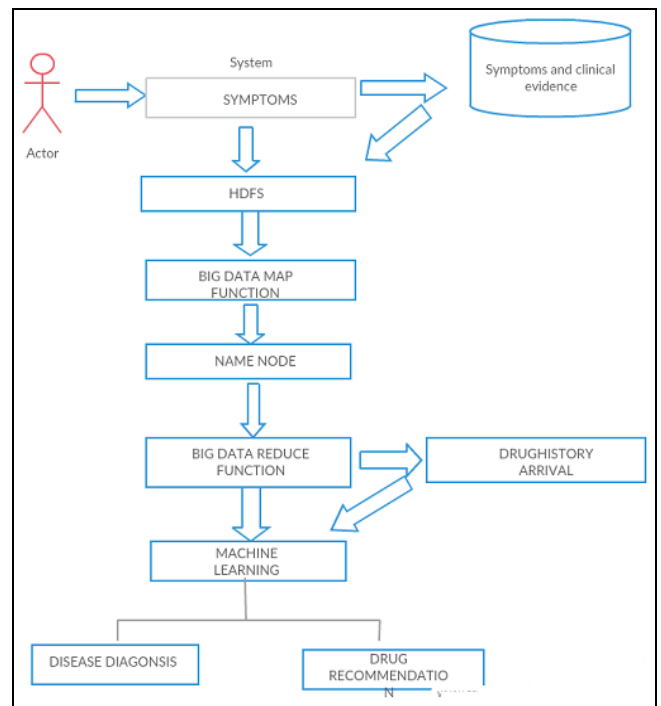
C. Use Case Diagram

Fig 2: - Clinical Decision Support system

Fig 2. shows details of the patients that are collected during the first visit of the patient for consultation. Evidence data includes information such as the treatments suggested, symptoms of disease, reaction to the prescribed drugs and laboratory results. Based on the patient symptoms and its details he or she will get the solutions as it retrieved data from the stored data or clinical data. The recorded data will be save through HDFS(Hadoop distributed file system).Where to handle these large data we are applying big data map function to it. Different data based on evidence of several disease will be distributed and stored to different nodes. Once the drug history(data about medicine, it's dosage and everything),through the technique of machine learning patient will get the solution about their disease easily and get proper course through online. It will save time and money. If the user is not able to get medicine easily through this proposed idea he or she will get as per their location it will show the nearest medicare centre where they will get proper medicine-it comes under GPS technology and facility like through maps.

VI.   METHODOLOGY

A. Using K-Means Method

Clustering is the process of partitioning a group of data points into a small number of clusters. K-means is a clustering method that aims to find the clusters that minimize the square of the distance from the data points to the cluster. This algorithm has been researched extensively and a detailed history of k-means.

- *Clustering approach*

According to clustering approaches can be broadly divided into two categories as follows: i) Hierarchical

clustering, and ii) Partitional clustering. On one hand, the family of hierarchical clustering algorithms attempt to build a hierarchy of clusters representing a nested grouping of objects and similarity level that change the grouping scope. In this way, clusters can be computed either in an agglomerative (bottom-up) fashion, or divisive (top-down) fashion. On the other hand, partitional clustering algorithms decompose data into a set of disjoint clusters Data is divided into K clusters satisfying that: i) Each cluster contains at least one point, and ii) Each point belongs to exactly one cluster. The K-Means algorithm is an example. We embraced the partitional clustering approach using the K-Means algorithm. The K-Means algorithm has been constantly considered as one of the simplest and most widely used clustering algorithms. The reasons behind the popularity of K-Means are robustly recognized in literature. Examples are the ease of implementation, simplicity, efficiency, and empirical success The K-Means clustering uses a simple iterative technique to group points in a dataset into clusters that similar characteristics. Initially, a number of clusters (K) is decided. The algorithm iteratively places data points into clusters by minimizing the within-cluster sum of squares..The algorithm converges on a solution when meeting one or more of these conditions:

a. The cluster assignments no longer change.

b. The specified number of iterations is completed.

- *Selected features*

The patients were clustered based on: i) Age, ii) LOS(length of stay), and iii) Time to surgery (TTS). We considered the numeric features only since the K-Means algorithm is originally applicable to numeric features only, such that a distance metric (e.g. Euclidean distance) can be used for measuring the similarity between data points. However, it is worth mentioning that there are some K-Means extensions that attempted to incorporate categorical features, such as the K-Modes algorithm.

B. *Using Map-Reduce Method*

Map Reduce is a programming model for processing huge volume of data in parallel. Map- Takes a single <key, value> pair, and produces zero or more new <key, value> pairs that may be of different type. Reduce-Works on the output of Map function and produces desired result. In the proposed model, Patient's Symptoms and the clinical evidences from the database are given as an input to the Map() function as given in Fig 3. Various symptoms across different diseases are mapped in the form of <key, values>pairs. Then symptoms and its most probable disease are identified by considering all clinical records available in the repository. The disease that has the high probability of all the symptoms specified by the patient is extracted. Subsequently, the best drug to treat the disease will be recommended by the system based on the success rate of drugs.
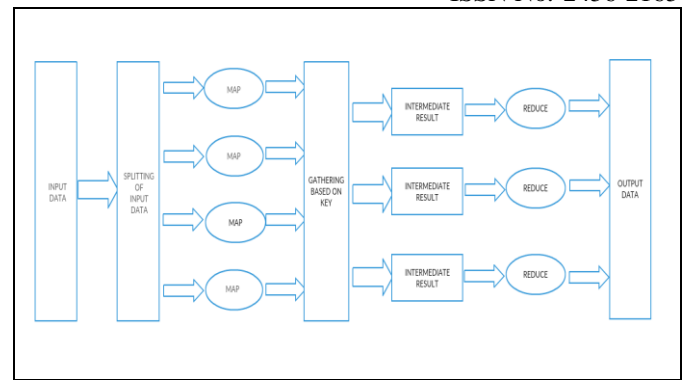

Fig 3:- Map Reduce Function

Input: Symptoms and Clinical Evidences
Output: Disease Diagnosis and Drug recommendation

C. *Using Cognitive Methods*

- *Trajectory Tracking*

Trajectory tracking is the macrocognitive process of imagining how unexpected events may affect practice and plans. Trajectory tracking is about getting ahead of the curve and projecting possible scenarios to be able to handle situations. In medicine, this process requires a functional perspective of how the disease will progress and what is expected based on what is observed. Clinical experts use this cognitive process of planning and replanning to prepare for unexpected events. This process provides a holistic overview of the patient's situation.

- *Perspective Taking*

The experts used group conformity to reduce the social pressure associated with antibiotic prescribing, primarily by engaging in perspective taking or simulating other minds. In this process, they imagine other practitioners' decision logic for a certain medical intervention or outcome.

VII. CONCLUSION

Proof based solution (EBM) is cognizant, particular, sensible utilization of current, best confirmations in settling on choices about treatment of individual patients. It's anything but a medical record-book with formulas, however its great application brings savvy and better social insurance. Its genuine object is that by the utilization of the most ideal confirmation specialist decides for his patient the most ideal arrangement, needing to give to him the ideal medicinal services in each angle. It is in like manner used to avoid huge stumbles all through treatment, and by this raises the idea of gave human administrations to the patient. In a broader setting, it can save the lives of our patients. Once the specialist pros the interest strategy and the usage of EBM he/she gets solid "accomplice " in their day by day respectable work. Since each physician must, in one way or another, seek valid information, we can say that the proper use of EBM saves doctors time and raises his level as well as the quality of provided medical services, and increases satisfaction of the health professionals. While EBM continues to be the guiding principle, clinicians should be aware of potential tainted results. In the future, big data is likely going to offer us a new aspect of EBM and arm us with more comprehensive data when we make our clinical

decisions. The system had analyzed the symptoms and diagnosed the exact disease and also the suitable drug is suggested for the diagnosed disease. This diagnosed disease and the best suitable drugs for a patient will aid the physician in analyzing and making decisions about the kind of treatment and medication for a patient with certain health conditions. Practice-based evidence may also be useful for the development of practice guidelines. Although the evidentiary support for individual decision steps in a guideline comes primarily from literature-based evidence, as discussed above, a guideline's process flow is usually constructed on the basis of expert opinion only. With more practice-based information on clinical processes and events, however, guideline developers may be able to improve the way they design process flows. Physician is able to get the required information as a second opinion before actual treatment starts. Thus the reliability of health care is assured.

REFERENCES

[1]. Omar El-Gayar, PremTimsina, Dakota State University, USA,"Opportunities for Business Intelligence and Big Data Analytics In Evidence Based Medicine", IEEE Trans, 2014, pp .749-757.978-1-4673-9745-2 ©2016 IEEE.

[2]. Kelley, A.S., Mcgarry, K., Fahle, S., Marshall, S.M.,Du, Q., and Skinner, J.S., "Out-of-Pocket Spending in theLast Five Years of Life", J Gen Intern Med, 28(2), 20145 pp.304-309.

[3]. Sankaranarayanan.S and PramanandaPerumal.T."Diabetic prognosis using Data Mining methods andtechniques" Proceedings of ICICA,Coimbatore, India,2014.

[4]. W. Raghupathi and V. Raghupathi, ``Big data analytics in healthcare: promise and potential,'' Health Inf. Sci. Syst., vol. 2, no. 1, pp. 1_10, 2014.

[5]. Lau BD, Haut ER. Practices to prevent venous thromboembolism: a brief review. BMJ Qual Safety2014;23:187-95.

[6]. Every-Palmer S, Howick J. How evidence-based medicine is failing due to biased trials and selective publication. J Eval Clin Pract 2014. May 12.

[7]. Duerden M, Avery T, Payne R. Polypharmacy and medicines optimization: making it safe and sound. King's Fund, 2015.

[8]. Llewelyn H, Ang HA, Lewis D, Al-Abdullah A. Oxford Handbook of Clinical Diagnosis. Oxford University Press, 2014.

[9]. Gierisch JM, Myers ER, Schmit KM, McCrory DC, Coeytaux RR, Crowley MJ, et al. Prioritization of patient-centered comparative effectiveness research for osteoarthritis. Ann Intern Med 2014. May 13.

[10]. Montori VM, Brito JP, Ting HH. Patient-centered and practical application of new high cholesterol guidelines to prevent cardiovascular disease. JAMA 2014;311:465-6.

[11]. West CP, Dyrbye LN, Rabatin JT, Call TG, Davidson JH, Multari A, et al. Intervention to promote physician well-being, job satisfaction, and professionalism: a randomized clinical trial. JAMA Intern Med2014;174:527-33.