Is Crowdsourcing Patient-Reported Outcomes the Future of Evidence-Based Medicine? A Case Study of Back Pain

Is crowdsourcing patient-reported outcomes the future of evidence-based medicine? A case study of back pain

Mor Peleg[1,2], Tiffany I. Leung[3], Manisha Desai[1], Michel Dumontier[1,4]

[1] Stanford Center for Biomedical Informatics Research, Stanford University, CA, USA
[2] Department of Information Systems, University of Haifa, Israel morpeleg@is.haifa.ac.il
[3] Faculty of Health, Medicine and Life Sciences, Maastricht University, The Netherlands
4 Institute of Data Science, Maastricht University, The Netherlands

Abstract. Evidence is lacking for patient-reported effectiveness of treatments for most medical conditions and specifically for lower back pain. In this paper, we examined a consumer-based social network that collects patients' treatment ratings as a potential source of evidence. Acknowledging the potential biases of this data set, we used propensity score matching and generalized linear regression to account for confounding variables. To evaluate validity, we compared results obtained by analyzing the patient reported data to results of evidence-based studies. Overall, there was agreement on the relationship between back pain and being obese. In addition, there was agreement about which treatments were effective or had no benefit. The patients' ratings also point to new evidence that postural modification treatment is effective and that surgery is harmful to a large proportion of patients.

1 Introduction

Lower back pain is a prevalent chronic condition affecting 39% of the population, which causes long-term disability and agony to patients, loss of work days and large healthcare costs [1]. Diagnosis and treatment is complicated by the fact that there is no clear association between pain and abnormalities detected by spine imaging [2]. Hence, many patients who undergo corrective surgery continue to have pain. Treatment options include spine surgery, injections, medications, psychological interventions, exercise, nutritional supplements, and lifestyle change and self-management approaches. Although many treatments exist, very few were shown to have more than moderate effectiveness at long-term pain reduction [3]. Clinical trials often employ small cohorts and cannot point to effective treatments. Some of them even have contradictory results. Furthermore, outcomes, especially patient-reported, are rarely systematically reported in electronic health records (EHRs).

In order to compare treatments by their effectiveness, objective measures need to be complemented with subjective patient-reported outcome measures (PROM). PROM are PROM are well established indicators of patients' global health [3]. Collecting PROM is a challenging but growing effort, involving clinicians, medical researchers and most importantly, patients/consumers. Efforts by medical care providers focus on collecting from patients, in a standardized way, the changes is their health state, (e.g., level of

pain, physical function, anxiety) [3]. On the other hand, collecting PROM from consumer-centric platforms (e.g., PatientsLikeMe, idonine.com, HealthOutcome.org) attract millions of patients and have obtained patients' treatment experiences from ten thousands of patient first-hand, including patients' treatment ratings, which are not collected in provider-centric EHRs.

Healthoutcome.org is a consumer health website that allows patients to report and share their treatment and health outcomes for most common orthopedic injuries and conditions. The site provides aggregated patients treatment outcome ratings as well as access to each patient review that includes patient information, treatment outcome rating and optional free text description. HealthOutcome has over 110,000 treatments ratings from over 15,000 patients, gathered in less than a year. A set of 38 treatment options are offered to lower back pain patients for rating, including a large number (26) of non-invasive/non-pharmacologic options and new treatment options.

Such non-invasive treatments are usually not documented at such granularity in EHR systems. Moreover, information about treatment outcomes is not available directly to patients. The primary limitation of HealthOutcome is that it does not currently collect PROM with a validated item-set; apart from treatment outcome ratings, patients indicate basic information about themselves, including their injury status (cured, in pain, or recovering), as well as their age category, gender, chronicity, and number of weekly hours of physical exercise. A further limitation is that the information entered is not inspected by clinicians to verify validity. Nonetheless, its importance is in providing transparent data about established and new treatments, while allowing treatments comparison by prevalence and crowd-sourced score, which can be filtered according to the characteristics of the reporting patients.

In light of the promise, but recognizing the limitations of such social networks as tools for evidence collection, our main research question is: **How can PROM among patients with low back pain improve our knowledge of effectiveness and harm of available treatments?** To answer this question we address the following objectives:

(1) Characterization of the HealthOutcome dataset' features and (2) its potential biases; (3) validation of associations with treatment and treatment effectiveness known from the literature or evidence-based studies; (4) Demonstration of the types of data analysis that can be done to compare treatments effectiveness; and (5) Reflection on the value and limitations of the crowdsourcing patient reported treatment outcome ratings as a source of evidence, and directions in which it can be improved.

2 Background

Different data sources and/or study designs may be used to estimate the effect of treatment on an outcome, with each approach having its own limitations. Randomized controlled studies collect evidence on treatment effectiveness by recruiting a homogenous cohort of patients and then randomly assigning patients to one treatment group or another. Because the group of patients is homogenous and most importantly, because subjects are randomized to treatment assignment, differences in outcomes can be attributed to the treatments, as such as design effectively controls for confounding. Alas, such traditional ways of collecting evidence have important shortcomings. First, most patients do not fit the study's inclusion or exclusion criteria, hence the evidence

is not applicable to them. Second, they are expensive and time-consuming to conduct. Usually small cohorts are recruited which limits the validity of evidence that can be generated. Additionally, some studies of intervention have issues with compliance in a randomized controlled setting. Consequently, most studies performed to compare effectiveness of back pain treatments do not provide conclusive evidence [3]. Alternatively, evidence of treatment effectiveness can be collected prospectively from medical records. Such observational designs may be desirable in their inclusiveness of patients and measures, but they pose threats to obtaining a causal effect of treatment due to the presence of confounding. Section 2.1 reviews some statistically-controlled methods to address such evidence collection. When treatment effectiveness cannot be objectively assessed by laboratory tests (i.e., pain medicine), PROM are collected from patients. We review provider- and consumer-based systems for collecting PROM, noting their differences.

2.1 Statistically-controlled methods to collect evidence prospectively

In recent years, researchers started using electronic medical records as a source of evidence for computing treatment effectiveness. However, Hersh et al. [4] note that "EHR data from clinical settings may be inaccurate, incomplete, transformed in ways that undermine their meaning, unrecoverable for research [e.g., found in textual notes], of insufficient granularity, and incompatible with research protocols" [i.e., treatment recommended as a balance of what is best for patient care and patient preferences]. Moreover, in observational prospective studies, where there is no randomization to intervention, confounding variables, such as demographics, medications at baseline, and medical conditions, may correlate with both the treatment and outcome [5]. Further, in systems that are based on users' decision to report, sampling bias may occur. For example, physicians may under-report adverse events of drugs that are already trusted vs. reporting for new drugs [6], or patients may decide not to rate treatments that they see as less important. Selection bias may also occur because patients had received certain treatment because it was indicated based on their demographic or disease-state, which are also correlated to the outcome being studied, such as adverse drug effects [7][6] or treatment ratings. One of the most popular methods to address confounding and issues with selection bias is to use propensity score matching [5] to account for confounders.

2.2 Provider systems for collecting PROM

The National Institutes of Health have assembled a task force on research standards for chronic low back pain [3]. This task force developed a minimal data set of 40 data items, 29 of which were taken from the PROM Information System (PROMIS) instrument. These items were recommended as offering the best trade-off of length with psychometric validity. The full item-set collects medical history, including chronicity, demographics, involvement in worker's compensation, work status, education, comorbidity, and previous treatment. Key self-report domains include pain intensity, pain interference, physical function, depression, sleep disturbance, and catastrophizing. Provider-centric implementations of PROMIS have been Implemented, such as Collaborative Health Outcomes Information Registry (CHOIR, https://choir.stanford.edu/) [8]. All patients with a pain diagnosis who visit clinics that have implemented

CHOIR are asked to complete the PROMIS questionnaire prior to each visit. When matched with EHR data, created by clinicians, which record the treatments that patients received, these records can be analyzed together to compare treatment effectiveness on individual and cohort levels.

2.3 Using crowdsourcing to find clinical evidence for treatment effectiveness

Unlike provider based medical records, patient social networks introduce sampling bias because not all patients seen by clinicians are active in social networks. In addition, their reports are not validated by clinicians during encounters to assess problems in understanding the semantics of questions asked, correctness and completeness.

Bove et al [9] validated the multiple sclerosis (MS) rating scale used in PatientsLikeMe.com by asking MS patients from a MS clinic to use the scale to rate the severity of their disease and compared it to the physician-provided scores recorded in their medical records. Having established the validity of the rating scale, they found small nonparametric correlations between BMI and the disease course of MS, adjusting for age, sex, race, disease duration, and disease type.

Nakamura et al. [10] compared clinicians' and patients' perspectives on the effectiveness of treatments for symptoms of amyotrophic lateral sclerosis by comparing data from a traditional survey study of clinicians with data from PatientsLikeMe. The perception of effectiveness for the five symptom-drug pairs that were studied differed. But due to the small number of patients' ratings that were available at the time of the study (20-66), statistical significance could not be evaluated. Nakamura et al. note the difference between the effectiveness provided by patients based on their direct personal experience versus that provided by clinicians, which is indirect, aggregated from their perception of experience of multiple patients but also more systematic as it draws from their clinical knowledge. It is worth noting that the symptoms studied by the authors (sialorrhea, spasticity and stiffness) can also be observed directly by clinicians.

3 Methods

3.1 Data collection and data set

Patients freely choose whether to post their reviews to HealthOutcome. They may remain anonymous or sign in. The web site is publicized by targeted Facebook ads, sent to adults who have posted content relating to orthopedic problems. The study was approved to review deidentified data by the Stanford University Human Subjects Research and Institutional Review Board (Protocol 40070). Data was obtained for patients with back pain who reported during 12/2008-12/2016. Two comma separated value (csv) files were obtained: one containing 5230 reviews by patients. Columns included: review ID, timestamp, user ID, injury Status (in pain, recovering, cured), age category (18-34, 35-54, 55+), gender, pain chronicity (<6M, 6-18M, >18M), hours of physical activity per week (0-4, 4-8, 8+), repeat injury?, weight, height, location (city, state), #surgeries, #treatments, textual review. From height and weight, we computed body mass index (BMI) category.

The second csv file contained 44,592 treatment ratings provided by patients in their reviews. The columns included review ID, treatment name, treatment rating. There are

five possible ratings: worsened, not improved, improved, almost cured, or cured. The two csv files were joined via a script written in Python 3 which extended the first csv file to contain 38 additional columns, one for each possible treatment, recording the treatment ratings provided in a review to each of the possible treatments.

3.2 Data analysis

We tested the following hypotheses:

1) Patients who respond to the website are not meaningfully different from the targeted patients in terms of age and gender; this was determined through a two-sided Chi-squared test comparing frequency distributions for age and gender.

2) PROM are internally consistent; We address this hypothesis by evaluating consistency of reporting by comparing reviews among patients who entered multiple reviews. This was determined by manual inspection of a random subset of 5% of these reviews, to see if a patient's demographic data and set of diagnoses and ratings did not change from one report to the next when they were provided within a six-month time period.

3) Those with high BMI have greater back pain; determined by linear regression.

4) Treatments' effectiveness, as determined in evidence-based studies, will match with ratings by patients; This was determined by comparing the literature-based effective treatments to patient-rated treatments with majority of ratings being improved, almost cured, or cured, which are not harmful; harmful treatments would be those where at least 10% patients ranked them as "worsening".

5) We hypothesize that postural modifications (PM) is more effective than spinal fusion surgery (SFS), and we hypothesize that PM is more effective than laminectomy; We addressed these hypotheses through two approaches. The first used generalized linear regression with a logit link (using R's Logit package) adjusting for potential confounders (adjusted model) and the second similarly utilized a logistic regression but with propensity matching (propensity score model). More specifically, for the adjusted model, we included indicators for treatments of interest in the regression: PM, SFS, or laminectomy. In addition, we included potentially confounding demographic variables of age group, gender, number of treatments, number of physical activity hours a week, chronicity and BMI; these parameters were chosen because they were shown to be predictive of pain status in a decision tree learning analysis, outside the scope of this paper. Regression for predicting treatment showed that spinal stenosis was a potentially confounding variable, hence this diagnosis was added to the demographics confounding variables for propensity score matching (using R's Matchit package). We report the odds ratio and confidence intervals for each analysis.

4 Results

4.1 Data characterization

Characteristics of the responding and the targeted patients are listed in Appendix A [11]. Of all responding patients, 43% of the targeted patients were 55 years or older and

80% were women. Responding patients had a larger proportion of the 55+ age group (72.7%, p-value <0.0001) and a smaller percentage of women (78.2%, p-Value <0.0001. The representation of older patients is in concordance with the literature. Most of the patients with back pain are in pain (57.9% in HealthOutcome vs. 52.9% of adults 65 or older [12]). Only 5.7% are cured (the rest are recovering). Accordingly, most of the patients' treatment ratings indicate no improvement (52.1%). 37.5% indicate improvement, 2.9% indicate that they are almost cured and only 1.3% say that they are cured. The percentage of reviews of worsening is 6.2%, higher than the total of cured and almost cured patients. This grim picture is consistent with the literature, showing that back pain is most often a chronic condition.

Missing values. The percentage of missing data are: gender 7.4%; age 12.2%; pain chronicity 24.1%; physical activity 22.3%; injury status 24.5%; weight/height 41.7%.

Data quality and consistency. 1% (27 of 2706) "In Pain" patients inconsistently provided treatment scores of "cured". 32% of patients provided non-anonymous reviews. 8.3% of the reports were by patients who each provided two or more reports. In a manual inspection of a random subset of 5% of these reports, we found that 5.6% of reports were inconsistent with respect to the patient's demographic data while 33% did not report the same set of diagnoses, treatments tried, or treatment ratings.

4.2 Consistency of relationships with those described in the literature

We evaluated whether relationships observed in our data set were consistent with those reported in the literature. Specifically, that high body mass is associated with an increased prevalence of low back pain [13]. Table 1 shows that the hypothesis that pain status is independent of BMI status can be rejected, supporting the hypothesis that obese patients are more in pain than others (linear regression; p-value=0.007; OR=0.17).

Table 1. Patient injury status with different BMI

BMI status	In Pain	Recovering	Cured
Underweight	25 (1.2%)	2 (0.5%)	3 (3.4%)
Normal	378 (18.6%)	101 (26.9%)	38 (43.2%)
Overweight	655 (32.2%)	145 (38.6%)	27 (30.7%)
Obese	975 (48.0%)	128 (34.0%)	20 (22.7)
Total	2033	376	88

Next, we compared patient opinions about treatment effectiveness to evidence based results. We first studied the distribution of treatment ratings. Table 2 shows select results. Complete results are in [11]. Treatments are ordered by prevalence. The mode is shown in bold. Treatments that worsen the state of at least 10% of patients are circled. Effective treatments (i.e., have ≥50% ratings in improved, almost cured, or cured and have <10% ratings of worsened) are shown in capitals. Not shown are treatments that were tried by fewer than 200 patients and ratings for broad classes of treatments – surgery and physical therapy (PT). The individual treatments under these categories are shown (e.g., PT includes TENS, stretching, heat, etc). The patients who were cured provided 365 treatment ratings of "cured". The treatments that received the highest number of "cured" ratings were strengthening exercises 41/365; postural modifications 37/365; and stretching 30/365. Table 3 compares the benefit of treatments according to their ratings by patient to results of evidence-based studies [14–16].

Table 2. Summary of treatment ratings

Patient ratings: Treatments	Worsened	Not improved	Improved	Almost cured	Cured	#patients tried treatment
NSAIDs	113	**1608**	946	49	9	2725
Cortisone Injection	151	**1308**	778	86	20	2343
REST	71	1088	**1102**	56	12	2329
Stretching	87	**1084**	991	78	40	2280
Strengthening Exercises	142	**1080**	868	64	52	2206
Chiropractor	171	**901**	661	69	39	1841
Epidural	101	**805**	492	70	9	1477
MASSAGE	48	662	**666**	44	19	1439
Acupuncture	23	**414**	181	29	4	651
SWIMMING	25	264	**299**	38	5	631
Spinal Fusion Surgery	126	**212**	167	37	18	560
Oral corticosteroids	16	**275**	184	22	3	500
Laminectomy Surgery	82	**171**	162	29	24	468
YOGA	14	132	**160**	22	17	345
POSTURAL MODIF.	8	118	**124**	22	39	311
Discectomy Surgery	33	85	**89**	12	11	230
All treatments	2189	**20229**	14571	1144	527	38660

Table 3. Comparison of treatment benefit: evidence-based vs. patient ratings

Patient ratings	Evidence from clinical trials: Effective	No Benefit	Harmful	No suffecient evidence
Effective	Massage Yoga Exercise (swimming)		Rest	***Postural modifications***
No Benefit	Acupuncture Spinal manipulation (chiropractor)	Steroid injection	Traction Inversion table	
Harmful				***Spinal Fusion*** ***Laminectomy*** ***Disectomy***
Not enough data	Functional restoration Interdisciplinary rehab Cognitive-behavioral	Prolotherapy	Home care Topical gel Dithermy	

4.3 Comparing treatment effectiveness

As expected, some attenuation in estimates of association were observed across models (Table 4). More specifically, the adjusted model had estimates that were closer to the null than the unadjusted model, and the propensity-score based model had estimates that were attenuated relative to the adjusted model. All models, however, provided evidence that PM was strongly associated with Cured status relative to SFS (Unadjusted OR=7.96, p-value < 0.001; Adjusted OR=6.61, p-values =0.014; and propensity score-based OR=6.52, p-value=0.025). Further, both the unadjusted and adjusted models provided evidence of an associated between PM vs Laminectomy and Cured status, whereas the propensity-score based method did not indicate a significant association (Unadjusted OR=10.03, p-value<0.001; Adjusted OR=5.16, p-value =0.029; and propensity score-based OR=5.08, p-value=0.065). In addition, results

from the propensity score-based model suggested that variables associated with Cured status include spinal stenosis when the diagnoses were considered, or #treatments, when they were not.

Table 4. Analysis of association of treatment options and patients having outcome of Cured

Treatment	N (before matching)	Odds Ratio (97.5% CI)	p-Value
Traditional unadjusted regression analysis			
Postural Modifications vs. Spinal fusion surgery	637	7.96 (4.10-17.03)	7.75e-09
Postural Modifications vs. Laminectomy	637	10.02(4.53-26.61)	2.09e-07
Traditional adjusted regression analysis			
PM vs. SFS considering demographics+DXs	355	6.61 (1.64-35.35)	0.01
PM vs. Lam considering demographics+Dxs	355	5.16 (1.30-26.91)	0.03
Propensity-score matched analysis			
PM vs. SFS considering demographics+Dxs	224 (236)	6.52 (1.40-40.83)	0.025
PM vs. Lam considering demographics+Dxs	224 (231)	5.08 (1.01-35.20)	0.065

5 Discussion

Patient crowdsourcing has been shown to provide large quantities of data. The quality of the data collection metrics, and the ability to validate the collected data, could be improved by collecting additional PROM, collecting data from wearable sensors about physical activity, and by linking the patient's reports to provider-based medical records. Even in its current state, the results suggest that patient-reported opinions of treatment effectiveness in a consumer social network are mostly consistent with published medical evidence. The most effective treatments were confirmed to be massage, yoga, and swimming exercises. In a small number of treatments, patients reported lower effectiveness than published literature suggests (acupuncture, spinal manipulation) or higher effectiveness (rest). The data also points to effective treatments that have not been studied in the evidence-based literature, including postural modifications, as well as provides evidence that all forms of surgery are considered as harmful by 14.4-22.5% of patients. These findings are in line with the recommendations of clinical guidelines [17] to delay surgery to later stages, in the absence of neurologic deficits. Surprisingly, 51% and 55% of patients found that stretching and strengthening exercises were not helpful. More details should be collected to evaluate this further. Generalizability to other domains and other designed platforms would also need evaluation.

Regression vs. Propensity score matching

Propensity score matching is considered a state of the art approach for handling confounding in observed studies, especially when there are ≤7 events per confounder [18]. But traditional multivariate regression models may also be appropriate in certain settings. Our analysis has shown (1) that the method of adjusting for confounding may produce disparate findings; Importantly, traditional unadjusted regression that does not account for confounding variables shows some results that are not replicated in analyses that accounts for confounding. Namely, that postural modification has better outcomes than laminectomy. In fact, laminectomy seemed to have a higher odds ratio vs. spinal fusion surgery, as compared to PM; and (2) that propensity matching developed to mitigate confounding can result in attenuated estimates even using the same

confounders in a traditional regression; traditional adjusted regression (considering demographics with/without diagnoses) showed that PM was superior to both SFS and laminectomy. However, these results were confirmed on the propensity-matched data set only for PM being superior to SFS; regression that considered demographics with diagnoses and was performed on the propensity-matched data did not show that PM had better outcomes than laminectomy.

Limitations and Future research

Incompleteness. The information collected directly from patients, using an easy-to-use interface, allows collection of a large volume of data quickly. However, incomplete data may result from the voluntary nature of reporting; we speculate that the high rate of missing weight and location data may seem too private to share. In addition, HealthOutcome's user interface has changed over time, so items that were added later (e.g., #surgeries) have more missing values.

Data quality and consistency. Inconsistency in reports by consumers is common. However, these are random measurement errors, which mostly just increase the variance of mean results. Evaluation with live subjects and corroboration with clinician-recorded data could estimate how well patients understand the items that they rate or indicate as being true. For example, are they aware of their diagnoses? Do they understand that reporting a value of zero or not reporting is not equivalent (e.g., #surgeries)? Do they consider long or short-term relief (e.g., rest)?

Generalizability. The large volume of data collected in consumer networks may help address the limitations discussed above, resulting in treatment ratings that could suggest evidence of effectiveness. However, the patient population in the consumer network does not represent all patients with back pain. Specifically, 80% of the users are women, probably reflecting the tendency of women to write posts on personal topics [19]. While the low number of back pain patients reported being cured (5.7%) could be attributed to patients' interest in reporting negative experience or to the severity of this disease, the former seems unlikely, considering that the percentage of patients who reported being cured of plantar fasciitis in HealthOutcome is much higher (27.8%) than back pain. Note however, that sampling bias is also present in provider-based PROM systems such as CHOIR, which collect the more severe patients who visit pain clinics. We thus suggest that the quantitative results would better be used qualitatively, pointing to potentially beneficial treatments. Conversely, the results pertaining to the high ratio of harmful surgeries do not distinguish patients who had the clinical indications for such treatments, who could benefit more from such treatment. In line with this, the fact that in HealthOutcome, spinal fusion surgery and laminectomy were more prevalent than postural modifications may indicate either that patients were referred to surgery before exhausting all non-invasive options or that a sampling bias was present.

Suggestions for improving data collection by HealthOutcome

The addition of the timing, frequency and duration of treatments, as well as the time until a review was provided, would be useful information to improve the analysis of back pain. The addition of PROMIS outcome measures could also generate a more valid assessment of outcome.

Acknowledgement

We thank Ofer Ben-Shachar for supplying the HealthOutcome data and thank him and Tobias Konitzer for the valuable discussions.

References

1. Hoy D, Bain C, William G, March L, Brooks P, Blyth F, et al. A systematic review of the global prevalence of low back pain. Arthritis Rheum. 2012;64:2028–37.
2. Chou R, Deyo RA, Jarvik JG. Appropriate use of lumbar imaging for evaluation of low back pain. Radiol Clin North Am. 2012;50:569–85.
3. Deyo RA, Dworkin SF, Amtmann D, et al. Report of the NIH Task Force on Research Standards for Chronic Low Back Pain. Int J Ther Massage Bodyw. 2015;8(3):16–33.
4. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research. Med Care. 2014;5:S30–7.
5. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. Br Med J. 2009;(338):b81.
6. Tatonett NP, Ye PP, Daneshjou R, Altman RB. Data-Driven Prediction of Drug Effects and Interactions. Sci Transl Med. 2013;4(125):1–26.
7. Harpaz R, DuMouchel W, Shah NH, et al. Novel Data Mining Methodologies for Adverse Drug Event Discovery and Analysis. Clin Pharmacol Ther. 2012;91(6):1010–21.
8. Bhandari RP, Feinstein AB, Huestis SE, et al. Pediatric-Collaborative Health Outcomes Information Registry (Peds-CHOIR): a learning health system to guide pediatric pain research and treatment. Pain. 2016;157(9):2033–44.
9. Bove R, Secor E, Healy B, et al. Evaluation of an online platform for multiple sclerosis research: patient description, validation of severity scale, and exploration of BMI effects on disease course. PLoS One. 2013;8(3):e59707.
10. Nakamura C, Bromberg M, Bhargava S, et al. Mining online social network data for biomedical research: a comparison of clinicians' and patients' perceptions about amyotrophic lateral sclerosis treatments. J Med Internet Res. 2012;14(3):e90.
11. Peleg M. Appendices. 2017. http://mis.haifa.ac.il/~morpeleg/PatientOutcomesAppend.html
12. Morone NE, Greco CM, Moore CG, et al. A Mind-Body Program for Older Adults With Chronic Low Back Pain: A Randomized Clinical Trial. JAMA Int Med. 2016;3:329-37
13. Heuch I, Hagen K, Heuch I, et al. The impact of body mass index on the prevalence of low back pain: the HUNT study. Spine (Phila Pa 1976). 2010;35(7):764–8.
14. Chou R, Huffman LH. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Soc. Ann Intern Med. 2007;147:492-504
15. Chou R, Atlas SJ, Stanos SP, Rosenquist RW. Nonsurgical interventional therapies for low back pain: a review of the evidence for an American Pain Soc. Spine 2009;34:1078-93
16. Chou R, Huffman LH. Medications for acute and chronic low back pain: a review of the evidence for an American Pain Soc. Ann Intern Med. 2007;147:505–14.
17. Institute for Clinical Systems Improvement. Adult Acute and Subacute Low Back Pain. Updated November 2012. 2012;
18. Biondi-Zoccai G, Romagnol E, Agostoni P, et al. Are propensity scores really superior to standard multivariable analysis? Contemp Clin Trials. 2011;32(5):731–40.
19. Wang YC, Burke M, Kraut RE. Gender, topic, and audience response: an analysis of user-generated content on facebook. SIGCHI Conf Hum Factors Comp Sys 2013:31-4