

# The Role of the Electronic Medical Record in the Assessment of Health Related Quality of Life

Serguei V.S. Pakhomov, PhD<sup>1</sup>, Nilay D. Shah, PhD<sup>2</sup>,  
Holly K. Van Houten<sup>2</sup>, Penny L. Hanson<sup>2</sup>,  
Steven A. Smith, MD<sup>2,3</sup>

<sup>1</sup> Pharmaceutical Care and Health Systems, University of Minnesota, Minneapolis, MN, USA ; <sup>2</sup> Health Care Policy & Research, Mayo College of Medicine, Rochester, MN, USA; <sup>3</sup> Endocrinology, Mayo College of Medicine, Rochester, MN, USA

## Abstract

*We applied a hybrid Natural Language Processing (NLP) and machine learning (ML) approach (NLP-ML) to assessment of health related quality of life (HRQOL). The approach uses text patterns extracted from HRQOL inventories and electronic medical records (EMR) as predictive features for training ML classifiers. On a cohort of 200 patients, our approach agreed with patient self-report (EQ5D) and manual audit of the EMR 65-74% of the time. In an independent cohort of 285 patients, we found no association of HRQOL (by EQ5D or NLP-ML) with quality measures of metabolic control (HbA1c, Blood Pressure, Lipids). In addition; while there was no association between patient self-report of HRQOL and cost of care, abnormalities in Usual Activities and Anxiety/Depression assessed by NLP-ML were 40-70% more likely to be associated with greater health care costs. Our method represents an efficient and scalable surrogate measure of HRQOL to predict healthcare spending in ambulatory diabetes patients.*

## Introduction

Numerous studies have demonstrated that the unstructured text of electronic medical records (EMRs) contains valuable information pertaining to patient health states as well as the process of healthcare delivery.<sup>1-9</sup> The objective of this study was to demonstrate the “meaningful use” of the EMR in determining patient functional health status. Our specific aim was to assess and compare the predictive validity and the utility of natural language processing and machine learning approaches to extracting physician observations as predictors of patient outcomes and healthcare resources utilization.

## Background

Patient’s functional status (FS) and health related quality of life (HRQOL) have been reported to be important predictors of intermediate<sup>10,11</sup> and long term patient outcomes.<sup>11-20</sup> However, these studies have been limited to using short term surveys or assessment of hospitalized patients<sup>10-16, 18-23</sup> and select outpatient populations<sup>24-26</sup> with more frequent problems with self care and usual activities.<sup>16, 19, 21-24</sup> These populations are not representative of the average individual seen long term in primary care. If measures of HRQOL are to permit planning clinical and health policy to close gaps in patient centric health achievement in chronic disease management<sup>27</sup>; there is a need for reliable and efficient methods of gathering this information and a demonstration of it’s value in predicting outcomes for patients seen in usual care settings.

## Materials and Methods

### Participants

Patients seen in primary care in the Mayo Clinic Health System outpatient clinics (n=454) completed postal mailings of the EuroQol5D (EQ5D) as part of the UNITED Planned Care Trial (UPC Cohort); a population based randomized controlled study assessing the value of shared care in the management of diabetes patients in primary care, [clinicaltrials.gov](http://clinicaltrials.gov): NCT00421850.<sup>28</sup> The EQ5D is a standardized questionnaire previously used in patients with diabetes to assess functional status in five domains; mobility, pain, self care, usual activities and anxiety/depression.<sup>29, 30</sup> The responses on the questionnaire ranged between 1 and 3 with 1 being the lowest value (no problems with functioning), 2 – (some problems with functioning) and 3 – (major problems with functioning).

An additional 200 patients not part of the UPC Cohort were included for validation studies (Validation Cohort) for the NLP-ML System. Both patient groups are representative of the six primary care family and internal medicine practices affiliated with Mayo Clinic, a large academic medical center in Rochester, Olmsted County, Minnesota, USA. The Mayo Foundation Institutional Review Board approved the study procedures and all patients participating gave written informed consent and research authorization.

### **Natural Language Processing and Machine Learning Approach**

We relied on a simple pattern-matching approach previously reported<sup>31</sup> for the extraction of predictive features used to training and automatic classifier (NLP-ML System) for the computerized determination of HRQOL from the text of electronic medical records. Machine learning algorithms applied to unstructured text typically require the following two steps: feature selection and feature extraction. We defined five sets of word patterns that were indicative of abnormal functional status in five domains of functioning: mobility, self-care, usual activities, pain, and depression using both top-down and bottom-up methods described in the following sections. The process by which these text patterns were defined (feature selection) is described in detail in the next two sections. Once the patterns have been defined, we encoded them as regular expressions using Perl programming language and applied these regular expressions to determine the presence and frequency of the features in EMR documents (feature extraction).

#### **Top-down (expert-knowledge) feature selection**

Because it was our intent to create an NLP-ML System that reflected documentation of functional status within the text of the medical record, we first completed a systematic review of published & unpublished inventories for assessing general and disease specific; i.e. diabetes and health status. Key words, phrases, and concepts that would identify textual references to FS were chosen and cataloged into one of five domains (mobility, usual activities, self-care, pain/discomfort, and anxiety/depression) based on their consistency with the International Classification of Functioning (ICF) construct<sup>32, 33</sup> as well as questions from previously validated instruments for the measure of functional status; Health Utilities Index Mark 2 (HUI2)<sup>34, 35</sup>, Health Utilities Index Mark 3 (HUI3)<sup>34-36</sup>, Quality of Well-Being Scale<sup>37, 38</sup>, Self-Administered V1.04 (QWB-SA)<sup>38</sup>, EuroQol5D<sup>39 40</sup>, SF-36v2<sup>41</sup>, the Patient Health Questionnaire (PHQ-9)<sup>42, 43</sup>, and a functional status assessment (HALex) questionnaire derived from the Behavioral Risk Factor Surveillance Survey: Center for Disease Control Telephone Administered Version)<sup>44-46</sup>. Using a modified Delphi approach we distributed these key words and concepts to domain experts in the assessment of functional status (rheumatology, epidemiology, endocrinology, health economics) until there were no further additions or clarification. Based on this review, a Measurement Manual was created (available upon request) that operationally defined for text auditors textual references to FS and would be sufficient documentation for the assessment of the five functional status domains.

#### **Bottom-up (data-driven) feature selection**

From the 454 UPC Cohort, we randomly selected 169 (37%) for review and annotation of their clinical notes for the 2 year trial period (July 2001- December 2003). Clinical notes sections documenting individual visits in the 2 year period from all 169 patients were first electronically retrieved and then randomly presented to the auditors (SS, PH) for independent manual annotation at the sentence level, using the General Architecture for Text Engineering (GATE 3.0).<sup>47</sup> The auditors electronically highlighted phrases in clinical notes indicative of functional status and assigned whether each highlighted portion of text was indicative of normal or abnormal status for the corresponding domain. In this way, review and annotation of the random selection was completed without regard to the integrity of all clinical notes for one patient. Following independent annotation and computation of inter-rater agreement; each document annotation was reviewed by the auditors together and, following discussion and consensus on disagreements, the final determination of functional status expressions was completed. These expressions were used as a source of additional keyword patterns for the construction of the NLP-ML System.

## Creation and Validation of the NLP System

We first used text queries of the designate clinical notes to extract a vocabulary of indicator phrases that signal evidence for FS assessment in order to identify linguistic patterns and bring lexical and morphological variants of medical terms to a standard form. We identified syntactic phrases representative of FS including noun phrases (e.g. “patient”), prepositional phrases (e.g. “with pain”), and adjective/adverb phrases (e.g. “very tired”) using two reference standards; 1) key words for functional status prediction, and 2) text annotated key words. After extracting all available phrases and their frequencies of occurrence in the records of patients with and without perfect health in each of the five FS domains (as well as if the condition was not assessed), information gain was measured for each of the phrases as an indication of its relevance to describing different health states and evidence for documentation. The measure of information gain indicates the discriminative power of a predictive feature with respect to a specific classification problem. Using features with high information gain values may improve classification accuracy, whereas using features with low information gain may add noise and result in poorer classification performance<sup>48</sup>. Each word or phrase in the vocabulary was treated as a potential predictor variable for a binary outcome classifier: positive or negative in reference to functional status. In addition, each clinical note was considered as an unordered list of predictor variables after stop words; e.g. “he”, “she”, “has”, “of”, were removed (“bag-of-words” representation<sup>49</sup>). We also applied limited semantic normalization (conceptual indexing) by using Metamap<sup>50</sup> that mapped free text of clinical reports to concept unique identifiers (CUIs) of the Unified Medical Language System Metathesaurus<sup>1</sup>.

Information gain value for each predictor was then computed and the words/concepts were ranked in the descending order of their information gain values.<sup>51</sup> Words/concepts with positive information gain were considered as potential candidates for inclusion in further search queries. These words combined with additional keywords and phrases as codified in the Measurement Manual were then used to construct natural language queries of the 5 FS domains and “Not Assessed”. Morphologic variants of the same word (e.g. move – moves – moving) were normalized by using Lexical Variant Generator <http://medlineplus.nlm.nih.gov/research/umls/meta4.html>. If a clinical note contained evidence for the assessment of a FS domain, then the text of the note was converted to a vector of predictive covariates. Support Vector Machine (SVM) learning algorithm (WEKA SMO implementation<sup>51</sup>) was then used to train a set of binary classifiers to determine if the FS component represented normal or abnormal status.<sup>52-56</sup> To train and validate the SVM algorithm, we represented each of the clinical note sections in terms of a set of predictive covariates.<sup>57</sup>

Ten-fold cross-validation strategy, using the WEKA data mining software package, was used to evaluate machine learning classifiers.<sup>49,51</sup> In addition, The NLP-ML System was tested on medical records for the 200 individuals in the Validation Cohort independent of the data used in training and cross-validation. These records were manually audited for the five functional status domains by SS and PH and each patient was classified as either “normal” or “abnormal” for each domain. In addition, each patient in this set filled out the EQ5D and patient responses were dichotomized to “normal” and “abnormal” categories using two as the cutoff value (> 2 – “abnormal”). All clinical notes for each patient in this cohort were also processed using the NLP system that made automated determination of “normal” vs. “abnormal” status for each domain. The resulting sets of responses from the auditors, patients and the NLP system were compared for agreement with each other.

### Clinical Outcomes

The clinical outcomes of interest in this study were the association of FS domains with quality performance measures and total cost of care for a 1 year period of observation (at some point during the study period spanning July 2001-December 2003). We used administrative data to estimate hospital and physician costs incurred by enrolled patients for one year prior to and after enrollment into the trial. A standardized, 2007 constant-dollar cost estimate was assigned to each service using the Medicare Part A and Part B classification system.<sup>58,59</sup> Specifically, Part A billed charges were adjusted by using hospital department cost-to-charge ratios and wage indexes, and Part B physician service costs were approximated by 2007 Medicare reimbursement rates.

The five domains of functional status for 285 UPC Cohort patients (Evaluation Cohort), not included in the process of pattern induction, were classified by the NLP System as present or absent and were used in regression models

---

<sup>1</sup> For this study we used an older Java implementation of Metamap (a.k.a. MMTx) as the newer Metamap system was not yet widely available at the time of the study.

including patient sex, age at diagnosis (years), duration of diabetes (years), and BMI as additional independent variables. The dependent performance variables for the logistic models were HbA1c < 7%, LDL-cholesterol < 2.6 mmol/l (100 mgm/dl), Blood Pressure < 130/80, and compliance with all three variables. Estimates for the United Kingdom Prospective Diabetes Study (UKPDS) 10-year risk for coronary heart disease (CHD)<sup>60</sup> and total costs were assessed using generalized linear models, specifying a gamma distribution and log link<sup>61, 62</sup>, each model using the same covariates as above. Odds ratio (95% confidence intervals) and parameter estimates are reported.

## Results

### Patient Demographics

Table 1 summarizes the demographic variables for the Validation and Evaluation Cohorts.

**Table 1 Patient Demographics for the Validation and Evaluation Cohorts**

Patient Demographics	Validation Cohort n=200		Evaluation Cohort n=285	
	Number (%)	Mean (standard deviation)	Number (%)	Mean (standard deviation)
% Male	97 (49)		139 (49)	
Age at diagnosis, years		48.6 (15.6)		55.1 (13.2)
Duration of diabetes, years		16.3 (11.9)		6.7 (7.4)
BMI kg/m <sup>2</sup>		33.2 (7.1)		33.1 (6.2)
Systolic BP mmHg		125 (15)		131 (16)
Diastolic BP mm Hg		69 (10)		72.6 (10.5)
HbA1c %		7.3 (1.1)		7.6 (1.6)
LDL Cholesterol mmol/l		2.2 (0.7)		2.7 (0.9)
UKPDS 10-year risk of CHD		25 (11)		21 (15)
Performance Metrics	Number (%)		Number (%)	
% HbA1c < 7 %	86 (43)		115 (40)	
% BP < 130/80 mmHg	113 (57)		115 (40)	
% LDL < 2.6 mmol/l	157 (79)		127 (45)	

The Evaluation Cohort was older at the time of diagnosis, had had diabetes for substantially less time, and except for LDL cholesterol levels less than 2.6 mmol/liter (100 mgm/dl), had similar metabolic control and 10 year estimate risk for coronary artery events as the Validation Cohort.

Table 2 summarizes the results of the comparison between the “normal” and “abnormal” classifications of functioning in five domains (mobility, self-care, usual activities, pain and depression) performed by the automated NLP-ML System, patient self-report on the EQ5D questionnaire and manual audit of the EMR for the Validation Cohort. The results indicate that the NLP-ML System output agreed on average 74% of the time with the manual audit results and approximately 65% with the patient self-report. The patient self-report and manual audit agreed approximately 71% of the time, while the auditors agreed with each other approximately 82% of the time. Thus, as expected, we observed better agreement between the NLP-ML system and the auditors than the patients. The distribution of “normal” and “abnormal” responses on EQ5D is typically skewed towards more prevalent “normal” responses, with the exception of pain domain. In our data, for example, in the Validation Cohort the percent of “normal” responses was 57.3% for mobility, 94.2% for self-care, 64.1% for usual activity, 66.7% for anxiety/depression, and 28.2% for pain.

The next question we addressed was whether patient’s responses or NLP-extracted functional status were better predictors of the outcomes.

**Table 2 Agreement between patient self-report, manual medical record audit and automatic NLP System for the Validation Cohort (n=200)**

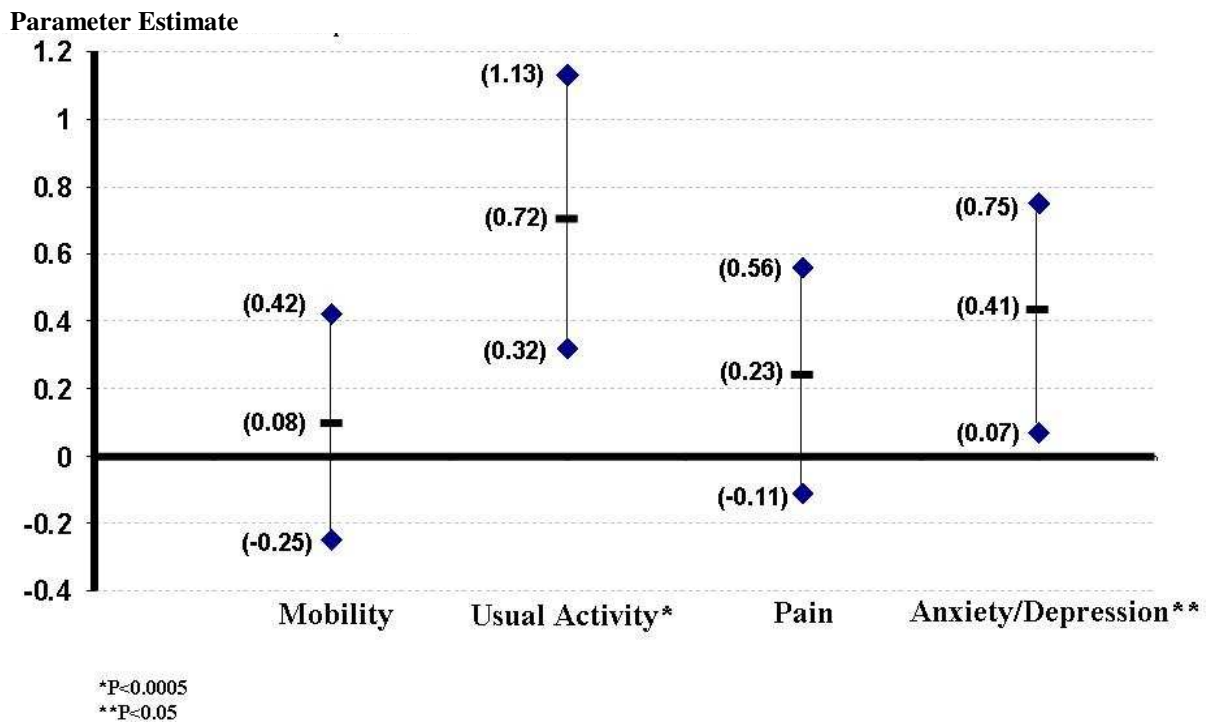
<b>Domain</b>	<b>Comparison</b>	<b>Kappa</b>	<b>% agreement</b>
<b>mobility</b>	SS vs PH*	0.4872	77.0
	Patient vs SS	0.2144	63.1
	Patient vs PH	0.2285	62.1
	Patient vs NLP-ML	0.0935	59.2
	SS vs NLP-ML	0.3655	79.0
	PH vs NLP-ML	0.2282	67.0
<b>self-care</b>	SS vs PH	0.5159	96.5
	Patient vs SS	0.0000	94.2
	Patient vs PH	0.1908	93.2
	Patient vs NLP-ML	0.1351	78.6
	SS vs NLP-ML	0.0927	81.5
	PH vs NLP-ML	0.1563	82.0
<b>usual activity</b>	SS vs PH	0.3337	75.5
	Patient vs SS	0.2127	65.0
	Patient vs PH	0.1825	65.0
	Patient vs NLP-ML	0.2485	66.0
	SS vs NLP-ML	0.3498	74.5
	PH vs NLP-ML	0.3179	74.0
<b>pain/discomfort</b>	SS vs PH	0.5714	80.5
	Patient vs SS	0.1328	66.0
	Patient vs PH	0.1027	64.1
	Patient vs NLP-ML	0.1746	63.1
	SS vs NLP-ML	0.3050	67.0
	PH vs NLP-ML	0.3886	70.5
<b>anxiety/depression</b>	SS vs PH	0.5552	81.0
	Patient vs SS	0.2969	70.6
	Patient vs PH	0.1654	63.7
	Patient vs NLP-ML	0.1081	56.9
	SS vs NLP-ML	0.3184	68.0
	PH vs NLP-ML	0.3748	70.0

\*SS, PH= auditors completing manual medical record audit for functional status

Patient= patient self report of functional status determined by EQ5D

NLP= automatic NLP System for functional status

We excluded from further analysis, abnormalities in the self care domain because this category had very low frequency by all assessments: patient self report, findings in manual review of the EMR, and the NLP-ML System. We found no association between functional status domains assessed by patient self report or NLP System and the performance measures; HbA1c, LDL cholesterol, BP, the three measures combined, or the 10 year risk for CHD (data not shown). While we did not find an association of patient self report of functional status and health care costs; abnormalities in each of the four FS domains determined by the NLP-ML System were associated with increased health care expenditure. Figure 1 includes the results of multivariate modeling and suggests that there is a significant impact of depression and usual activity, as encoded in free text of the EMR and extracted with NLP, on health care costs.



**Figure 1.** Parameter estimates and 95% confidence intervals around the estimates for EQ5D domains in a multivariate model predictive of health care costs as the outcome relative to individuals without functional status problems.

## Discussion

In this study we have demonstrated the use of NLP and the EMR in determining an efficient and scalable surrogate measure of HRQOL that predicts health care utilization and cost of care for ambulatory patients with diabetes. Our results show that two of the five EQ5D domains of functioning (i.e., “usual activity” and “anxiety/depression”) are significantly associated health care utilization. Our preliminary efforts suggest that automated processing of the unstructured text of clinical reports has the potential to automate the collection of a rich set of information concerning measures of HRQOL that could inform quality improvement interventions. These findings suggest that our approach may be a scalable and efficient population strategy that uses NLP for applying possible interventions targeting usual activity and depression. Such targeted and evidence-supported interventions (as opposed to, for example, blindly targeting improvements in A1c) could result in more efficient health care utilization through reduced costs.

Our study has strengths and weaknesses regarding the measure and use of HRQOL. Similar to other reports using manual chart abstraction, we have found that there are variable levels of agreement using NLP derived measure of HRQOL compared to patient interview and self report. While it is intuitive that the more concerning a disability is to the patient, the more likely it would be a subject of communication and documentation in the medical record, past studies have been inconsistent and have found that health care providers accurately<sup>23,24,26</sup>, over<sup>26,63</sup>, and under estimate<sup>13,19,21,23,26,63-66</sup> patient HRQOL. We may speculate that when a patient goes to see a physician he/she is more likely to have some concerns on his/her mind vs. to have a general examination. The patient response may then obviously be biased in one situation or the other. In addition, providers themselves may not accurately assess functional status and thus one would not expect NLP to be any better than medical record documentation. Furthermore, providers might over-document exaggerated patient report in the encounter vs. survey when they are anticipating adequate documentation for justification in care delivery or the visit. Based on more affected populations, medical record documentation has had reported sensitivity (35-88%) and specificity (64-82%) for predicting problems with usual and self care<sup>19, 21, 24, 26</sup> with similar but less frequently reported predictive value for domains of anxiety/depression and pain.<sup>24, 26</sup> The purpose of our study was not to resolve the issue of disagreements between clinical documentation and patient self-report, but to compare medical record documentation, as well as patient self report for HRQOL to health care expenditure. Despite differences in self

report and use of the medical record to assess functional status, our findings support prior observations that medical record data appears to be able to serve as a predictor for increased health care utilization.<sup>19</sup>

The features used to train the machine learning classifiers came partly from a manual process that was necessary to provide the fidelity of the algorithm; however, it may introduce bias. Therefore, the results of the current study would need to be replicated in another documentation system and another population.

## Conclusion

We believe this is the first report of the use of the medical record to determine the value of a measure of HRQOL and its association with the cost of care for an ambulatory population. In addition, we have described a process that can easily be translated to other care settings using the EMR, permitting an assessment of HRQOL that is not dependent on integration into an already time constrained clinical encounter in primary care.

## References

1. Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. *Journal of the American Medical Informatics Association* 2000;7:593-604.
2. Friedman C. A broad-coverage natural language processing system. *Proceedings / AMIA 2000;Annual Symposium.*:270-4.
3. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. *Journal of the American Medical Informatics Association* 1994;1:161-74.
4. Hazlehurst B, Sittig DF, Stevens VJ, et al. Natural language processing in the electronic medical record: assessing clinician adherence to tobacco treatment guidelines. *American Journal of Preventive Medicine* 2005;29:434-9.
5. Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *Journal of the American Medical Informatics Association* 2005;12:448-57.
6. Mendonca EA, Haas J, Shagina L, Larson E, Friedman C. Extracting information on pneumonia in infants using natural language processing of radiology reports. *Journal of Biomedical Informatics* 2005;38:314-21.
7. Pakhomov SS, Hemingway H, Weston SA, Jacobsen SJ, Rodeheffer R, Roger VL. Epidemiology of angina pectoris: role of natural language processing of the medical record. *American Heart Journal* 2007;153:666-73.
8. Persell SD, Wright JM, Thompson JA, Kmetik KS, Baker DW. Assessing the validity of national quality measures for coronary artery disease using an electronic health record. *Archives of Internal Medicine* 2006;166:2272-7.
9. Pakhomov S, Shah N, Hanson P, Balasubramaniam S, Smith SA. Automatic quality of life prediction using electronic medical records. *AMIA 2008;Annual Symposium Proceedings/AMIA Symposium.*:545-9.
10. Davis RB, Iezzoni LI, Phillips RS, Reiley P, Coffman GA, Safran C. Predicting in-hospital mortality. The importance of functional status information. *Medical Care* 1995;33:906-21.
11. Covinsky KE, Justice AC, Rosenthal GE, Palmer RM, Landefeld CS. Measuring prognosis and case mix in hospitalized elders. The importance of functional status.[see comment]. *Journal of General Internal Medicine* 1997;12:203-8.
12. Winograd CH, Gerety MB, Chung M, Goldstein MK, Dominguez F, Jr., Vallone R. Screening for frailty: criteria and predictors of outcomes. *Journal of the American Geriatrics Society* 1991;39:778-84.
13. Pinholt EM, Kroenke K, Hanley JF, Kussman MJ, Twyman PL, Carpenter JL. Functional assessment of the elderly. A comparison of standard instruments with clinical judgment. *Archives of Internal Medicine* 1987;147:484-8.
14. Narain P, Rubenstein LZ, Wieland GD, et al. Predictors of immediate and 6-month outcomes in hospitalized elderly patients. The importance of functional status. *Journal of the American Geriatrics Society* 1988;36:775-83.
15. Inouye SK, Wagner DR, Acampora D, et al. A predictive index for functional decline in hospitalized elderly medical patients.[see comment]. *Journal of General Internal Medicine* 1993;8:645-52.
16. Inouye SK, Peduzzi PN, Robison JT, Hughes JS, Horwitz RI, Concato J. Importance of functional measures in predicting mortality among

- older hospitalized patients. *JAMA* 1998;279:1187-93.
17. Hoenig H, Hoff J, McIntyre L, Branch LG. The self-reported functional measure: Predictive validity for health care utilization in multiple sclerosis and spinal cord injury. *Archives of Physical Medicine & Rehabilitation* 2001;82:613-8.
  18. Daley J, Jencks S, Draper D, Lenhart G, Thomas N, Walker J. Predicting hospital-associated mortality for Medicare patients. A method for patients with stroke, pneumonia, acute myocardial infarction, and congestive heart failure. *JAMA* 1988;260:3617-24.
  19. Burns RB, Moskowitz MA, Ash A, Kane RL, Finch MD, Bak SM. Self-report versus medical record functional status. *Medical Care* 1992;30:MS85-95.
  20. Brorsson B, Asberg KH. Katz index of independence in ADL. Reliability and validity in short-term care. *Scandinavian Journal of Rehabilitation Medicine* 1984;16:125-32.
  21. Rodriguez-Molinero A, Lopez-Dieguez M, Tabuenca AI, de la Cruz JJ, Banegas JR. Functional assessment of older patients in the emergency department: comparison between standard instruments, medical records and physicians' perceptions. *BMC Geriatrics* 2006;6:13.
  22. Elam JT, Graney MJ, Beaver T, el Derwi D, Applegate WB, Miller ST. Comparison of subjective ratings of function with observed functional ability of frail older persons. *American Journal of Public Health* 1991;81:1127-30.
  23. Bogardus ST, Jr., Towle V, Williams CS, Desai MM, Inouye SK. What does the medical record reveal about functional status? A comparison of medical record and interview data.[see comment]. *Journal of General Internal Medicine* 2001;16:728-36.
  24. Velikova G, Wright P, Smith AB, et al. Self-reported quality of life of individual cancer patients: concordance of results with disease course and medical records.[erratum appears in *J Clin Oncol* 2001 Oct 15;19(20):4091]. *Journal of Clinical Oncology* 2001;19:2064-73.
  25. Rubenstein LV, Calkins DR, Young RT, et al. Improving patient function: a randomized trial of functional disability screening. *Annals of Internal Medicine* 1989;111:836-42.
  26. Newell S, Sanson-Fisher RW, Girgis A, Bonaventura A. How well do medical oncologists' perceptions reflect their patients' reported physical and psychosocial problems? Data from a survey of five oncologists. *Cancer* 1998;83:1640-51.
  27. Kerr EA, Smith DM, Hogan MM, et al. Building a better quality measure: are some patients with 'poor quality' actually getting good care? *Medical Care* 2003;41:1173-82.
  28. Smith SA, Shah ND, Bryant SC, et al. Chronic care model and shared care in diabetes: randomized trial of an electronic decision support system. *Mayo Clinic Proceedings* 2008;83:747-57.
  29. Anonymous. Quality of life in type 2 diabetic patients is affected by complications but not by intensive policies to improve blood glucose or blood pressure control (UKPDS 37). U.K. Prospective Diabetes Study Group. *Diabetes Care* 1999;22:1125-36.
  30. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Annals of Medicine* 2001;33:337-43.
  31. Pakhomov SVS, Hanson PL, Bjornsen SS, Smith SA. Automatic classification of foot examination findings using clinical notes and machine learning. *Journal of the American Medical Informatics Association* 2008;15:198-202.
  32. Ustun TB, Chatterji S, Kostansjek N, Bickenbach J. WHO's ICF and functional status information in health records. *Health Care Financing Review* 2003;24:77-88.
  33. ICF. Online version: ICF-English. In: World Health Organization; 2007.
  34. Luo N, Johnson JA, Shaw JW, Feeny D, Coons SJ. Self-reported health status of the general adult U.S. population as assessed by the EQ-5D and Health Utilities Index. *Medical Care* 2005;43:1078-86.
  35. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Annals of Medicine* 2001;33:375-84.
  36. Maddigan SL, Feeny DH, Majumdar SR, Farris KB, Johnson JA. Health Utilities Index mark 3 demonstrated construct validity in a population-based sample with type 2 diabetes. *Journal of Clinical Epidemiology* 2006;59:472-7.
  37. Hanmer J, Lawrence WF, Anderson JP, Kaplan RM, Fryback DG. Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores.[see comment]. *Medical Decision Making* 2006;26:391-400.
  38. Pyne JM, Sieber WJ, David K, Kaplan RM, Hyman Rapaport M, Keith Williams D. Use of the quality of well-being self-administered version (QWB-SA) in assessing health-related quality of life in depressed patients. *Journal of Affective Disorders* 2003;76:237-47.
  39. Tsuchiya A, Brazier J, Roberts J. Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. *Journal of Health Economics* 2006;25:334-46.



40. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics* 2004;13:873-84.
41. Glasziou P, Alexander J, Beller E, Clarke P, Group AC. Which health-related quality of life score? A comparison of alternative utility measures in patients with Type 2 diabetes in the ADVANCE trial. *Health & Quality of Life Outcomes* 2007;5:21.
42. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *Journal of General Internal Medicine* 2007;22:1596-602.
43. Sullivan MD, Anderson RT, Aron D, et al. Health-related quality of life and cost-effectiveness components of the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial: rationale and design. *American Journal of Cardiology* 2007;99:90i-102i.
44. Erickson P. Evaluation of a population-based measure of quality of life: the Health and Activity Limitation Index (HALex). *Quality of Life Research* 1998;7:101-14.
45. Livingston EH, Ko CY. Use of the health and activities limitation index as a measure of quality of life in obesity. *Obesity Research* 2002;10:824-32.
46. Yabroff KR, McNeel TS, Waldron WR, et al. Health limitations and quality of life associated with cancer and other chronic diseases by phase of care. *Medical Care* 2007;45:629-37.
47. Cunningham H. GATE, a General Architecture for Text Engineering. *Computers and the Humanities* 2002;36:1572-8412.
48. Bowyer AF. Quantitative information of specific diagnostic tests. *Journal of Medical Systems* 1998;22:3-13.
49. Manning C, Shutze H. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press; 1999.
50. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc* 2010;17:229-36.
51. Whitten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed. San Francisco: Elsevier; 2005.
52. Aphinyanaphongs Y, Tsamardinos I, Statnikov A, Hardin D, Aliferis CF. Text categorization models for high-quality article retrieval in internal medicine. *Journal of the American Medical Informatics Association* 2005;12:207-16.
53. Cohen AM. An effective general purpose approach for automated biomedical document classification. *AMIA Annu Symp Proc* 2006:161-5.
54. Hissa M, Pahikkala T, Suominen H, et al. Towards automated classification of intensive care nursing narratives. *Stud Health Technol Inform* 2006;124:789-94.
55. Joshi M, Pedersen T, Chute CG. A comparative study of supervised learning as applied to acronym expansion in clinical reports. *AMIA Annu Symp Proc* 2006:399-403.
56. Xu H, Markatou M, Dimova R, Liu H, Friedman C. Machine learning and word sense disambiguation in the biomedical domain: design and evaluation issues. *BMC Bioinformatics* 2006;7:334.
57. Pakhomov S, Weston SA, Jacobsen SJ, Chute CG, Meverden R, Roger VL. Electronic Medical Records for Clinical Research: Application to the Identification of Heart Failure. *American Journal of Managed Care* 2007;13:281-8.
58. Gold M, Siegel JE, Russell LB, Weinstein M. *Cost-Effectiveness in Health and Medicine*. New York: Oxford Univ Press; 1996.
59. Lave JR, Pashos CL, Anderson GF, et al. Costing medical care: using Medicare administrative data. *Medical Care* 1994;32:JS77-89.
60. Clarke PM, Gray AM, Briggs A, et al. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia* 2004;47:1747-59.
61. Blough DK, Ramsey SD. Using generalized linear models to assess medical care costs. *Health Services and Outcomes Research Methodology* 2000;1:185-202.
62. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *Journal of Health Economics* 2001;20:461-94.
63. Calkins DR, Rubenstein LV, Cleary PD, et al. Failure of physicians to recognize functional disability in ambulatory patients. [see comment]. *Annals of Internal Medicine* 1991;114:451-4.
64. Stephens RJ, Hopwood P, Girling DJ, Machin D. Randomized trials with quality of life endpoints: are doctors' ratings of patients' physical symptoms interchangeable with patients' self-ratings? *Quality of Life Research* 1997;6:225-36.
65. Rubenstein LZ, Schairer C, Wieland GD, Kane R. Systematic biases in functional status assessment of elderly adults: effects of different data sources. *Journal of Gerontology* 1984;39:686-91.
66. Grossman SA, Sheidler VR, Swedeen K, Mucenski J, Piantadosi S. Correlation of patient and caregiver ratings of cancer pain. *Journal of Pain & Symptom Management* 1991;6:53-7.