Assessing Information Congruence of Documented Cardiovascular Disease between Electronic Dental and Medical Records

Jay Patel BDS, MS[1,4], Danielle Mowery MS, PhD[2,3], Anand Krishnan MS[1], Thankam Thyvalikakath DMD, MDS, PhD[1,4,5]
[1]Indiana University School of Dentistry, Indianapolis, IN, [2]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, [3]Informatics, Decision-Enhancement, and Analytic Sciences Center (IDEAS 2.0), Veterans Affairs Salt Lake City Health Care System, Salt Lake City, UT, [4]Department of Bio-Health Informatics, IUPUI School of Informatics and Computing, Indianapolis, IN; [5]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN

Abstract

Dentists are more often treating patients with Cardiovascular Diseases (CVD) in their clinics; therefore, dentists may need to alter treatment plans in the presence of CVD. However, it's unclear to what extent patient-reported CVD information is accurately captured in Electronic Dental Records (EDRs). In this pilot study, we aimed to measure the reliability of patient-reported CVD conditions in EDRs. We assessed information congruence by comparing patients' self-reported dental histories to their original diagnosis assigned by their medical providers in the Electronic Medical Record (EMR). To enable this comparison, we encoded patients CVD information from the free-text data of EDRs into a structured format using natural language processing (NLP). Overall, our NLP approach achieved promising performance extracting patients' CVD-related information. We observed disagreement between self-reported EDR data and physician-diagnosed EMR data.

Introduction

Thanks to advances in medical research and treatments, people are living longer and leading better lives than ever before [1,2]. As of 2015, 25% of the United States population (approximately 80 million adults) fall within the 65 years, and older age group; this age group is expected to increase by 50% in 2040 [1]. Despite living a longer life, most of these individuals live with multiple comorbid chronic conditions such as diabetes, osteoporosis, cardiovascular diseases (CVDs), and dental diseases such as periodontal disease [2,3]. Among these medical conditions, CVD is one of the most common comorbidities experienced by this population [4]. They are more often treated in the dental clinics [5)]. Unsurprisingly, CVDs are the most frequent conditions reported by patients seeking dental care [5,6].

As a result, dental clinicians need to carefully review and monitor patients' current CVD status applying more advanced clinical decision making and treatment planning as some dental procedures can be invasive [7,8]. For example, if CVDs are not well controlled before the dental procedure, then it could lead to serious consequences and adverse events [7,8] as discussed at the *World Workshop on Oral Medicine VI: Controversies regarding dental management of medically complex patients: assessment of current recommendations* [9]. Typically, CVD information provided to dental clinicians are reported by the patient. It's unclear to what extent the information is accurate and represents the patients' actual CVD conditions [10,11]. Because self-reported conditions can vary based on the patient's' age, gender, education level, cognitive function, income level, and the presence of a number of chronic diseases [12–15].

In the existing literature, only one 1991 study by Levy et al. characterized information congruence by comparing dental patients' self-reported conditions with the information gathered from their medical providers [16]. They compared medical histories reported by dental patients and their physicians. Authors found that patients and physicians reported conditions not reported by the other indicating neither sources can be leveraged exclusively for reliable information [10]. However, the authors compared conditions from the patients' medical histories and rather than from the patients' diagnosis codes. It is unclear to what extent the information provided by physicians was accurate. Additionally, the authors have not included details regarding individual CVD concepts such as myocardial infarction, hypertension, and coronary artery disease, as they mainly focused on general conditions (central nervous system, cardiovascular, gastrointestinal, renal, pulmonary). Moreover, this study was done more than two decades ago, and patient demographics have changed since the last two decades [17].

With the increased use of electronic dental records (EDRs) and electronic medical records (EMRs) to document patient care during the last decade, we now have access to clinical information electronically[18,19]. The reliability of patient-reported CVD information could be measured by comparing CVD information present in their EDR with those in their EMR[10]. In EMRs, patients' CVD are documented by their primary care providers using diagnostic codes i.e., International Classification of Diseases (ICD) codes[20]. Through a partnership between Indiana University School of Dentistry and Regenstrief Institute, we have the opportunity to utilize Indiana Network for Patient Care (INPC) data. The INPC data is the country's oldest, largest, and most comprehensive regional health information exchange including clinical data from over 140 hospitals, local laboratories, imaging centers, and a few large-scale practices. INPC is a rich source of medical information, which is widely used for research and for point of care purposes. Indiana University School of Dentistry (IUSD) patient's dental data has been linked to their medical data through the INPC data source[21],allowing us to understand information congruence between EDR and EMR data.

For example, in the EMR, a patient's CVD information is stored in a structured format utilizing ICD diagnostic codes, e.g., ICD code 411.1 represents "intermediate coronary artery syndrome," and ICD code 433.1 represents "occlusion and stenosis of carotid artery." In the EDR, a patient's CVD information is stored in a free-text patient medical history field e.g., "mini heart attack," and "artery is half blocked" because diagnostic code usage in dentistry is not widely adopted. Structured codes from the EMR and unstructured data from EDR make it challenging to compare information congruence. Therefore, to understand information congruence between record sources, we applied natural language processing (NLP), an approach to structure information from free texts into a computable and comparable data format. NLP has been shown to accurately structure CVD-related concepts such as smoking [22], obesity [23], medications [24], and risk factors [25] from clinical notes in the EMR and can support this study. Although the NLP methods applied to this study are not novel, to our knowledge, our study is the first to assess information congruence (or loss) of CVD information between the EDR and EMR an underserved area of research.

Therefore, our pilot study consisted of two objectives: (1) *encode patients' CVD-related information automatically from the EDR using NLP* and (2) *compare CVD information congruence between linked EMR and EDR data*.

Methods

To achieve these objectives, we completed the following six steps. *First*, we created guidelines to annotate patient's CVD status based on the type of CVD condition, CVD procedure, other CVD attributes. *Second*, we developed a reference standard of manually annotated unique patient CVD histories. *Third*, we split this reference standard into training and testing datasets. *Fourth*, we trained a regular expression-based NLP system to extract CVD information (CVD extractor). *Fifth*, we tested the performance of the CVD extractor. *Lastly*, we assessed CVD information congruence by comparing CVD concepts extracted from the EDR to the patient's matched EMR diagnostic codes.

Extracting and preprocessing EDR data

We obtained the de-identified EDRs (axiUm: Exan Corporation, Vancouver, BC, Canada) for 254 patients who underwent a comprehensive oral examination between January 1, 2011, and January 1, 2012, at the Indiana University School of Dentistry (IUSD). We matched our dental patients EDR with their EMR from INPC data generated from January 1, 2006, through January 1, 2012. We included those patients who have answered at least one CVD-related question in the medical history form (either presence or absence of the disease). We excluded those patients who did not provide any information and left all the fields blank in the medical history form.

Generating and characterizing the reference standard

We created annotation guidelines for manually annotating patients' CVD information that typically dentists seek during patient care. CVD-related information addressed by our annotation schema included CVD concepts of *conditions, procedures,* and *medications*, as well as CVD concept-related attributes of *severity*, *time*, and *experiencer* (**Table 1**). Using these guidelines, two researchers independently annotated 254 unique patients' dental histories. Additionally, condition, procedure, and medication concepts were mapped to the Unified Medical Language System (UMLS) dictionary[26], a collection of standard biomedical vocabularies that permit standardization of the annotation process and enable interoperability between the computer systems. We annotated patient dental histories using an annotation tool called extensible Human Oracle Suite of Tools (eHOST)[27,28]. eHOST permits assigning annotated text-based concepts to Concept Unique Identifiers (CUIs) from the UMLS. The CUI unifies variably written phrases into a single, unique, structured, and meaningful concept. For instance, in our CVD histories, the condition concept "Heart Murmur" is written with variations e.g., " H.M.", "Murmur", "Mur". While annotating these records, we assigned all these variations to one unique CUI, which is "C0018808" for Heart murmur in the UMLS dictionary.

The overall agreement between the two annotators was 0.90 (Cohen's Kappa value), which indicated an excellent agreement. Disagreements between the annotators were resolved through discussion and consensus. The finalized annotations were considered as the reference standard.

Table 1. Summary of annotation guidelines for manually annotating patient's CVD–related information

CVD concepts & attributes	**Description and annotation guidelines**	**Examples of mapped literal text matches from EDRs**
CVD condition concept	This concept describes CVD-related signs, symptoms, and diagnoses; each are assigned a UMLS CUI.	C0018808: "heart murmur.", C0018790: "cardiac arrest 3 yrs ago", C0027051: "Pat. had an MI"
CVD procedure concept	This concept describes CVD-related procedures including diagnostic, preventative, and administered interventions; each are assigned a UMLS CUI.	C0010055: "heart attack surgery.", C0013516: "echocardiogram done."
CVD medication concept	This concept describes CVD-related preventative medications; each are assigned a UMLS CUI.	C0699129: "Patient takes coumadin.", C0019134: "Patient used heparin."
CVD negation attribute	This attribute describes the presence or absence of a CVD concept and is subcategorized into: 1) Affirmed: CVD concepts that are affirmed 2) Negated: CVD concepts that are negated 3) Possible: CVD concepts that are ambivalent or uncertain	1) Affirmed: "patient was diagnosed with myocardial infraction last week." 2) Negated: "diagnosis came negative and patient did not have heart attack" 3) Possible: "patient may have some artery plaque, but the diagnosis is not yet done.
CVD severity attribute	This attribute describes the severity (extent, degree, or amount) of CVD condition and is subcategorized into: 1) Mild: CVD concepts that are mild or low 2) Moderate: CVD concepts that are moderate 3) Severe: CVD concepts that are severe or high	1) Mild: "Patient had mild heart attack in 2002, but now he is doing fine." 2) Moderate: "mod myocardial infarc" 3) Severe: "Patient had severe heart attack and went for surgery in 2010."
CVD temporal attribute	This attribute represents the time of diagnosed CVD condition or procedures/medications performed for CVD and is subcategorized into: 1) Historical: CVD concepts occurring three years before the EDR date 2) Current: CVD concepts occurring within three years of EDR date 3) Not particular/conditional/future: CVD concepts that didn't necessarily occur at a particular time or occur conditional to other conceptual events	1) Historical: "Patient had heart cauterization performed 13 years ago" 2) Current: "heart attack in 2016" 3) Not particular/conditional/future: "patient mentioned that he had echocardiography done but cannot recall when that was done"
CVD experiencer attribute	This attribute represents the experiencer of the condition or procedure or receiver of the medication and is subcategorized into: 1) Patient: the current individual receiving care 2) Other: any of patient's family members.	1) Patient: "Patient had a heart attack in 2002." 2) Other: "patient's mother had heart murmur."

Evaluating the NLP-based CVD extractor performance

To encode each CVD concept and attribute type, we developed a simple CVD extractor that uses regular expressions derived from the training dataset and synonyms from the UMLS to encode CVD concepts and attributes embedded in free-text into a computable structured format, similar to the NegEx algorithm[29]. Next, we applied this program on the testing dataset. We evaluated the performance of the NLP-based, CVD extractor by calculating their precision, recall, and F-measure described by Hripcsak et al.[30] (**Table 2)**.

Table 2. Formulas to calculate precision, recall, and F-measure to evaluate CVD Extractor's performance

Computer program evaluation measures	**Formulas**
Precision	true positive / (true positive + false positive)
Recall	true positive / (true positive + false negative)
F-measure	2 * (precision * recall) / (precision + recall)

Assessing CVD information congruence between EDR and EMR data

Next, we aimed to assess the information congruence of each patient's affirmed CVD conditions between EDR and EMR data. Specifically, for each patient affirmed CVD condition extracted and validated from the EMR, we mapped the CUI to the associated ICD-9 diagnostic code. Then determined if the intermediate EDR ICD-9 diagnostic code occurred in among diagnostic codes within the patient's corresponding EMR. We assessed agreement between these two sources using the Cohen's Kappa statistical test[31] **(Eq. 1**). Cohen's kappa is a statistical measure of agreement that is calculated based on expected vs. observed values and frequencies.

(Eq. 1) $K=(P_0-P_e)/(1-P_e)$

Here, P_0 is the observed percent agreement; P_e is the expected percent agreement. Kappa values range between 0 and 1, with numbers closer to -1 indicating high disagreement and values closer to 1 indicating high agreement.

Results

We annotated CVD-related information from EDR history data, developed an NLP-based CVD extractor to structure CVD-related concepts and attributes into a computable format, and then assessed the information congruence of patient-reported CVDs between this EDR-based data and EMR-based diagnostic data.

Generating and characterizing the reference standard

Our reference standard data consisted of a total of 80 CVD conditions, 70 CVD procedures, and 19 medication concepts. **Table 3** depicts the most common CVD types in our population. We found 171 (67.3%) patients had at least one CVD condition among 254 patients. Additionally, we observed a total of 128 (50.4%) patients with one documented procedure performed out of 254. The most common procedures performed in our patients were cardiac stress test (18%), placement of stent (18%), echocardiogram (14%), and coronary artery bypass surgery (6%).

Table 3. Frequency distributions of concepts and attributes found in 254 patients' EDRs.

Cardiovascular Disease (CVD) concepts & attributes	**Occurrence in dental population**	**Most common CVD types**
CVD condition concept	80 CVD conditions	1) Heart murmur: 28 (16.4%) 2) Myocardial infarction: 19 (11.1%) 3) Coronary Artery Disease: 12 (7%)
CVD procedure concept	70 CVD procedures	1) Cardiac stress test: 23 (18%) 2) Placement of stent: 23 (18%) 3) Echocardiography: 18 (14%) 4) Coronary Artery Bypass Surgery (6%)
CVD medication concept	19 CVD medication concepts	1) Atorvastatin: 8 (38%)
CVD temporal attribute	1) Current: 35 2) Historical: 128 3) Future condition not particular: 13	Not Applicable
CVD negation attribute	1) Negation: 189 2) Affirmed: 50 3) Possible: 02	Not Applicable
CVD severity attribute	1) Mild: 7 2) Moderate: 3 3) Severe: 5	Not Applicable
CVD experiencer attributes	1) Patient: 249 2) Other: 3	Not Applicable

Evaluating the NLP-based CVD extractor performance

In **Table 4,** the CVD extractor achieved an overall F-measure of 85.42% for extracting and encoding various CVD concepts and attributes. The CVD extractor achieved high F-scores for extracting attributes of severity, experiencer, and negation. We suspect this excellent performance could be due to similar patterns of writing and fewer variations. For instance, most of the clinicians described the severity of patients CVD conditions using fairly standardized scales, e.g., levels of "low" "moderate", and "high" as well as intensities of "mild" and "severe". Additionally, fewer histories described family histories of CVD resulting in low error and high accuracy. In contrast, the CVD extractor performed moderately at extracting CVD-related conditions and procedures. We hypothesize this is due to large lexical variations e.g., acronyms, abbreviations, and misspellings used to represent these concepts e.g.,. "Heart murmur" represented as "HM", "mur", and "heart mumr". Additionally, we missed condition and procedural concepts that were not present in the training dataset or UMLS dictionary suggesting the need for more sophisticated vocabulary expansion methods, e.g., word embeddings[32].

Table 4. Performance of the CVD Extractor for encoding CVD condition concepts, CVD procedural concepts, CVD negation attributes, CVD experiencer attributes, CVD temporality attributes, and CVD severity attributes.

CVD concepts and attributes	**F-score**	**Precision**	**Recall**
Overall	85.42%	98.50%	76.70%
CVD condition concept	78.32%	100.00%	64.36%
CVD procedural concept	76.39%	98.31%	62.50%
CVD negation attribute	90.81%	94.38%	87.50%
CVD experiencer attribute	100.00%	100.00%	100.00%
CVD temporality attribute	76.42%	98.31%	62.50%
CVD severity attribute	90.55%	100.00%	83.33%

Assessing CVD information congruence between EDR and EMR data

When comparing CVD information between the EDR and EMR, we observed Cohen's Kappa value of -0.4 indicating disagreement. As described in **Table 5**, we found much inconsistency and wide variations between CVD information recorded in EDR and EMR. We observed a total of 288 unique CVD conditions related concepts between the EDR and EMR. 18% (52 of 288) of CVD condition related concepts were solely recorded in EDR; conversely, 72% (208 of 288) of CVD concepts solely recorded in patients matched EMR. In both the EDR and EMR data, we identified 28 present CVD concepts. Of the 28 shared CVD conditions, chest pain and hyperlipidemia were the most frequently reported in both EDR and EMR data. Of the 208 unshared concepts identified from EMR, hypertension and cardiomyopathy were the most commonly recorded concepts within EMR. Unique CVD concepts such as cardiomyopathy, chronic atrial fibrillation, pericardial disease, carotid artery syndrome, atrial flutter, chronic venous insufficiency were less frequent in the EDR, but more frequent in the EMR. In contrast, myocardial infarction, heart murmur, and coronary artery disease were more frequent in the EDR, but least frequent in the EMR.

Table 5. Comparison between CVD condition concepts found and not found in the EDR and EMR

CVD concepts	**Not found in EDR**	**Found in EDR**	**Total**
Not found in EMR	0	52	52
Found in EMR	208	28	236
Total	208	80	288

Discussion

For this pilot study, we demonstrated the feasibility of applying NLP to patient dental histories to extract CVD information for the purpose of understanding information congruence (in this case, information loss) between medicine and dentistry health records, an important and underserved area of research. We learned several important lessons from this initial investigation: 1) *CVD information can be accurately extracted from dental histories* and 2) *CVD information congruence is poor between EDR and EMR data.*

CVD information can be accurately extracted from dental histories

Our CVD extractor achieved excellent F-measure for extracting CVD attributes of temporal, negation, experiencer and severity. However, it achieved moderate F-measure for extracting CVD conditions and procedures, primarily due to moderate recall. This finding indicates the need for additional vocabulary expansion efforts and perhaps more sophisticated NLP methods. In the NLP literature, several NLP tools have been developed to extract CVD-related (heart disease) risk factors from clinical notes, particularly as part of the 2014 i2b2/UTHealth NLP Challenge.[36, 37–45] For instance, Khalifa et al. developed an NLP pipeline by adapting components from cTAKES and Textractor.[37] They achieved an F-measure of 87%; comparably, we achieved an F-measure of 85%. Similarly, Yang et al. have developed a hybrid model combining rules and machine learning components in their pipeline achieving an F-measure of 89%.[38] These studies demonstrate the range of more complex NLP approaches to extracting CVD-related variables. However, additional comparisons between the i2b2/UTHealth NLP Challenge task and our task are difficult to make. For example, the challenge task focused on a few CVD-related risk factors i.e., smoking and diabetes; whereas, we have addressed a variety of other CVD condition concepts e.g. myocardial infarction, heart murmur, coronary artery disease, rheumatic fever, congestive heart disease, mitral valve stenosis, and many more unique CVD procedural and medication concepts. Conversely, our task is far more constrained than the challenge task. Specifically, in the EDR, patients' CVD information is documented as a fairly succinct free-text field requiring minimal text processing; whereas, in the EMR, CVD information is dispersed throughout the entire clinical note requiring more extensive text processing. Our EDR-based CVD extractor achieved promising initial results demonstrating the feasibility of using a simple NLP approach to extracting a broad range of CVD concept from dental histories that appears moderately comparable to EMR-based tools addressing a similar task.

These promising results suggest that CVD extractor could be accurately applied to a much larger EDR dataset to advance our knowledge about CVD and its impact with various dental and systemic diseases, treatments, and prognosis. This approach may also help in determining the reliability of using EDR data for research purposes, which has been questionable in current format.

CVD information congruence is poor between EDR and EMR data

We observed high disagreement between the information recorded in the EDR compared with EMR data demonstrating low information congruence. Surprisingly, we found less commonly shared CVD conditions in EDR and EMR. For example, myocardial infarction was mostly documented in the EDR, but rarely in EMR. We suspect that this could be because physicians only document patients' myocardial infarction if diagnosable from clinical observations. Conversely, because EDR CVD information is self-reported, patients may write what they think they experienced. For instance, one history in EDR stated, "I had a mini heart attack". Cardiomyopathy, chronic atrial fibrillation, pericardial disease, carotid artery syndrome, atrial flutter, and chronic venous insufficiency were infrequently reported in EDR. We hypothesize that patients may not remember which CVD condition they had, understand the differences and relationships between each specific CVD conditions, or provide sufficient detailed clinical descriptions for dental practitioners to record each CVD condition precisely. Therefore, this information may not be reported reliably and subsequently documented in the self-reported dental history forms. CVD conditions such as myocardial infarction, stroke, and coronary artery disease are more prevalent in the general population. As a result, patients might be more predisposed or likely to hearing these conditions, understanding what they mean, and tend to report them. Okura et al. identified evidence to support this hypothesis, identifying that self-reported diseases (diabetes, myocardial infarction, and stroke) were frequently reported with high accuracy (90%) by patients[33]. Unfortunately, the extent of patients' knowledge about less common CVD conditions such as cardiomyopathy, chronic atrial fibrillation, pericardial disease, carotid artery syndrome, atrial flutter, chronic venous insufficiency is not well studied in the existing literature. Finally, patient's may not be aware that CVD conditions that are not well controlled before the dental procedure could lead to serious consequences and adverse events[7,8]. Therefore, patients may not feel this is pertinent information to disclosure with their dental providers. This presents an important opportunity for dentists to educate patients about the importance of disclosing and/or updating CVD information at each dental visit prior to receiving clinical care.

Compared with general medicine, our information congruence results contradict what is known in the literature, For instance, Merkin et al. observed variable agreement between end-stage renal disease patients and physicians who reported for diabetes (Kappa statistic k =0.93) and myocardial infarction (k=0.79), and chronic obstructive pulmonary disease (k=0.20) [13]. Similarly, Engstad et al. observed high sensitivity also known as recall (80%) and excellent specificity (99%) of self-reported stroke between affected patient populations and their providers [34]. While in our study, we report high disagreement between patients' reported dental histories and physicians diagnosis.

Compared with dental medicine, our information congruence results moderately confirm what is known in the literature with similar study design. For example, McDaniel et al. assessed patient willingness to reveal health history information, sexually transmitted diseases, substance abuse, HIV infections, and tuberculosis collected from history forms within the EDR. They observed that a significant number of patients provide inaccurate or incomplete information to questions routinely asked on these history forms [35]. Similarly, Levy et al. observed moderate agreement between patients and physicians reported general medical conditions when comparing collected medical and dental histories. Both patients and physicians reported systemic conditions not reported by the other suggesting that neither patients nor physicians can be relied as the sole source of information. Similarly, we observed that myocardial infarction was frequently recorded in self-reported the EDR but not in the EMR, while hypertension was frequently recorded in self-reported EMR, but not in EDR [10].

Limitations and Future Work

There are several limitations to our study. First, the CVD extractor may not generalize well beyond our institutional data because writing patterns may vary across different institutions, however, we believe many of our terms representing CVD concepts could be leveraged as a starting point for others seeking to extract this information from their own dental histories. In this study, we did not consider the effect of confounders on self-reporting these CVD conditions. In the next phase, we will examine the effect of confounders such as age, gender, and race on self-reporting CVD conditions. We also plan to test our CVD extractor using histories from other institutions. Given the success of this NLP method, we will extend the system to structure patients' smoking status documented within dental histories from IUSD. We aim to apply both tools to understand the prevalence of CVD and smoking status in our population, determine relationships between CVD and smoking status with dental diseases such as dental caries, and periodontal disease, among other research endeavors. In this study, we only measured the reliability of CVD conditions, not CVD procedure and CVD medication information. In future, we will use standard terminologies and vocabularies to assess information congruence between our annotated CVD procedures e.g., current procedural codes to find the reliability of CVD procedure recorded in EDR. Additionally, we will also find the reliability of CVD medication reported in EDR by comparing it with patients' prescription data.

Conclusion

Based on our initial study results, we conclude that it is feasible to extract patients CVD information automatically from EDRs utilizing NLP. We also believe there are important and unique opportunities to bridge the chasm of information loss for CVDs between the EDR and the EMR. This is demonstrated by the observation that patients were good at reporting prevalent CVD conditions such as myocardial infarction, and coronary artery disease; however, they did not report less prevalent CVD such as cardiomyopathy, chronic atrial fibrillation, and pericardial disease. Future studies are warranted to determine how dentists can be informed about patients' true CVD status in order to enable shared decision before starting treatment plan.

Acknowledgments

We would like to thank Ms. Wang Jiachen and Mr.Craig Eberhardt for generating dataset for this study, Mr. Anand Kulanthaivel reviewing the manuscript, and Mr. Andres Mantilla for formatting references. This study was approved by the Institute Review Board (IRB: 1604548307) and funded by Dr. Thyvalikakath's start-up funds from the Indiana University School of Dentistry.

References

1. Eke PI, Wei L, Borgnakke WS, Thornton-Evans G, Zhang X, Lu H, et al. Periodontitis prevalence in adults ≥ 65 years of age, in the USA. Vol. 72, Periodontology 2000. 2016. p. 76–95.
2. Wahid A, Chaudhry S, Ehsan A, Butt S, Ali Khan A, Khan AA. Bidirectional Relationship between Chronic Kidney Disease & Periodontal Disease. Pakistan J Med Sci. Professional Medical Publications; 2013 Jan;29(1):211–5.
3. Daar AS, Singer PA, Persad DL, Pramming SK, Matthews DR, Beaglehole R, et al. Grand challenges in chronic non-communicable diseases. Vol. 450, Nature. 2007. p. 494–6.
4. Go A, Mozaffarian D, VL R, EJ B, Berry J, Borden W, et al. Statistical Fact Sheet 2013 Update Nutrition & Cardiovascular Diseases Nutrition & CVD - 2013 Statistical Fact Sheet. Circulation. 2013;127:e6–245.
5. Dwivedi S, Sharma N, Sharma V. Dental considerations in cardiovascular patients: A practical perspective.

Vol. 69, Indian Heart Journal. Elsevier; 2017. p. 423–4.

6. Sanchez P, Everett B, Salamonson Y, Ajwani S, Bhole S, Bishop J, et al. Oral health and cardiovascular care: Perceptions of people with cardiovascular disease. Aalto-Setala K, editor. PLoS One. Public Library of Science; 2017 Jul 20;12(7):e0181189.
7. Burgess J. Dental Management in the Medically Compromised Patient Overview, Diabetes, Drug Reactions. Medscape. 2015.
8. Ettinger RL. Treatment planning concepts for the ageing patient. Aust Dent J. 2015 Mar 1;60(S1):71–85.
9. Napeñas JJ, Kujan O, Arduino PG, Sukumar S, Galvin S, Baričević M, et al. World Workshop on Oral Medicine VI: Controversies regarding dental management of medically complex patients: Assessment of current recommendations. In: Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology. 2015. p. 207–26.
10. Levy SM, Jakobsen JR. A comparison of medical histories reported by dental patients and their physicians. Spec Care Dent. 1991;11(1):26–31.
11. Radfar L, Suresh L. Medical profile of a dental school patient population. J Dent Educ. 2007 May;71(5):682–6.
12. Goldman N, Lin IF, Weinstein M, Lin YH. Evaluating the quality of self-reports of hypertension and diabetes. J Clin Epidemiol. 2003 Feb;56(2):148–54.
13. Merkin SS, Cavanaugh K, Longenecker JC, Fink NE, Levey AS, Powe NR. Agreement of self-reported comorbid conditions with medical and physician reports varied by disease among end-stage renal disease patients. J Clin Epidemiol. NIH Public Access; 2007 Jun;60(6):634–42.
14. Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. J Clin Epidemiol. 2004 Oct;57(10):1096–103.
15. Hansen H, Schäfer I, Schön G, Riedel-Heller S, Gensichen J, Weyerer S, et al. Agreement between self-reported and general practitioner-reported chronic conditions among multimorbid patients in primary care - Results of the MultiCare Cohort Study. BMC Fam Pract. BioMed Central; 2014 Mar 1;15(1):39.
16. Chen K, Li Z. How does information congruence influence diagnosis performance? Ergonomics. 2015 Jun 3;58(6):924–34.
17. Cohn D, Caumont A. 10 demographic trends that are shaping the U.S. and the world | Pew Research Center [Internet]. pewresearch.org. 2016 [cited 2018 Mar 7]. Available from: http://www.pewresearch.org/fact-tank/2016/03/31/10-demographic-trends-that-are-shaping-the-u-s-and-the-world/
18. Song M, Liu K, Abromitis R, Schleyer TL. Reusing electronic patient data for dental clinical research: A review of current status. Vol. 41, Journal of Dentistry. 2013. p. 1148–63.
19. Patel J, Siddiqui Z, Krishnan A, Thyvalikakath T. Identifying patients' smoking status from electronic dental records data. In: Studies in Health Technology and Informatics. 2017. p. 1281.
20. M.Kang&apos;ethe S, W. Wagacha P. Extracting Diagnosis Patterns in Electronic Medical Records using Association Rule Mining. Int J Comput Appl. 2014 Dec 18;108(15):19–26.
21. Siddiqui Z, Wang Y, Makkad P, Thyvalikakath T. Characterizing restorative dental treatments of sjögren's syndrome patients using electronic dental records data. In: Studies in Health Technology and Informatics. 2017. p. 1166–9.
22. Savova GK, Ogren P V, Duffy PH, Buntrock JD, Chute CG. Mayo Clinic NLP System for Patient Smoking Status Identification. J Am Med Informatics Assoc. 2008;15(1):25–8.
23. Kuo T-T, Rao P, Maehara C, Doan S, Chaparro JD, Day ME, et al. Ensembles of NLP Tools for Data Element Extraction from Clinical Notes. AMIA Annu Symp proceedings; 2016;2016:1880–9.
24. Halgrim SR, Xia F, Solti I, Cadag E, Uzuner Ö. A cascade of classifiers for extracting medication information from discharge summaries. J Biomed Semantics. BioMed Central; 2011;2(3):S2.
25. Stubbs A, Kotfila C, Uzuner Ö. Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1. J Biomed Inform. 2015 Dec;58:S11–9.
26. Kleinsorge R, Tilley C, Willis J. Unified Medical Language System (UMLS). Encycl Libr Inf Sci. U.S. National Library of Medicine; 2002;369–78.
27. Tahsin T, Beard R, Rivera R, Lauder R, Wallstrom G, Scotch M, et al. Natural language processing methods for enhancing geographic metadata for phylogeography of zoonotic viruses. AMIA Jt Summits Transl Sci Proc; 2014:102–11.
28. Mowery DL, Jordan P, Wiebe J, Harkema H, Dowling J, Chapman WW. Semantic annotation of clinical events for generating a problem list. AMIA Annu Symp proceedings; 2013;2013:1032–41.
29. University of Utah/Veteran Affairs NLP ToolFinder website - ConText/NegEx Algorithm [Internet]. [cited

2018 Mar 7]. Available from: http://toolfinder.chpc.utah.edu/content/contextnegex

30. Hripcsak G, Rothschild AS. Agreement, the F-measure, and reliability in information retrieval. J Am Med Informatics Assoc. American Medical Informatics Association; 2005;12(3):296–8.
31. Wiwanitkit V. Authorship for leading journals: Issue on ethical problems. Vol. 206, American Journal of Roentgenology. Croatian Society for Medical Biochemistry and Laboratory Medicine; 2016. p. W61.
32. Velupillai S, Mowery DL, Conway M, Hurdle J, Kious B. Vocabulary Development To Support Information Extraction of Substance Abuse from Psychiatry Notes. BioNLP. Stroudsburg, PA, USA: Association for Computational Linguistics; 2016;92–101.
33. Longtin Y, Sax H, Leape LL, Sheridan SE, Donaldson L, Pittet D. Patient participation: Current knowledge and applicability to patient safety. Vol. 85, Mayo Clinic Proceedings. Mayo Foundation; 2010. p. 53–62.
34. Engstad T, Bønaa KH, Viitanen M. Validity of Self-Reported Stroke : The Tromsø Study. Stroke. 2000;31(7):1602–7.
35. McDaniel TF, Miller D, Jones R, Davis M. Assessing patient willingness to reveal health history information. J Am Dent Assoc. Elsevier; 1995 Mar 1;126(3):375–9.
36. Afzal N, Sohn S, Abram S, Liu H, Kullo IJ, Arruda-Olson AM. Identifying peripheral arterial disease cases using natural language processing of clinical notes. In: 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). IEEE; 2016. p. 126–31.
37. Khalifa A, Meystre S. Adapting existing natural language processing resources for cardiovascular risk factors identification in clinical notes. J Biomed Inform. Academic Press; 2015 Dec 1;58:S128–32.
38. Yang H, Garibaldi JM. A hybrid model for automatic identification of risk factors for heart disease. J Biomed Inform. NIH Public Access; 2015 Dec;58 Suppl(Suppl):S171-82.
39. Liao KP, Ananthakrishnan AN, Kumar V, Xia Z, Cagan A, Gainer VS, et al. Methods to Develop an Electronic Medical Record Phenotype Algorithm to Compare the Risk of Coronary Artery Disease across 3 Chronic Disease Cohorts. Bamias G, editor. PLoS One. 2015 Aug 24;10(8):e0136651.
40. Meystre SM, Kim Y, Gobbel GT, Matheny ME, Redd A, Bray BE, et al. Congestive heart failure information extraction framework for automated treatment performance measures assessment. J Am Med Informatics Assoc. 2016 Jul 12;24(e1):ocw097.
41. Miranda E, Irwansyah E, Amelga AY, Maribondang MM, Salim M. Detection of Cardiovascular Disease Risk’s Level for Adults Using Naive Bayes Classifier. Healthc Inform Res. Korean Society of Medical Informatics; 2016 Jul;22(3):196–205.
42. Pakhomov SSV, Hemingway H, Weston SA, Jacobsen SJ, Rodeheffer R, Roger VL. Epidemiology of angina pectoris: Role of natural language processing of the medical record. Am Heart J. 2007 Apr;153(4):666–73.
43. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. In: 2008 IEEE/ACS International Conference on Computer Systems and Applications. IEEE; 2008. p. 108–15.
44. Srinivas K, Kavihta B, Govrdhan RA, Jagtial K. Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. Int J Comput Sci Eng. 2010;2(2):250–5.
45. Urbain J. Mining heart disease risk factors in clinical text with named entity recognition and distributional semantic models. J Biomed Inform. 2015 Dec;58:S143–9.