

Yumeng Liu. A Natural Language Processing Approach to Analyzing Consultation Visits for Treatment Decision Making for Localized Prostate Cancer. A Master's Paper for the M.S. in I.S degree. April, 2018. 36 pages. Advisor: Fei Yu

This quantitative study investigates the information seeking and giving behaviors of doctors, prostate cancer patients and their family members during consultation visits. The dataset was a transcript of audio-recordings of real-time treatment consultations collected from 171 consultation visits in a clinical trial. For text analysis purposes, a question mark in a transcript sentence was used to signal the information seeking behavior while a period was used to signal information giving behavior. Two counting methods were adopted to investigate the effect of a decision aid intervention that was created to facilitate the prostate cancer patient consultation. Data process, cleaning and analysis were performed by Python, which shows the decision aid intervention promoted the patient's information seeking and giving behavior during the consultations. Topic analysis of specific target population showed different communication styles from Caucasian and African America population. What's more, the results contribute to the decision aid for patients and training on communications in the clinical visits.

Headings:

Natural Language Processing

Topic Analysis

Decision Aid

A NATURAL LANGUAGE PROCESSING APPROACH TO ANALYZING
CONSULTATION VISITS FOR TREATMENT DECISION MAKING FOR
LOCALIZED PROSTATE CANCER

by
Yumeng Liu

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2018

Approved by

Fei Yu

Table of Contents

Table of Contents	1
Introduction	2
Literature Review	4
2.1 The decision making challenges that prostate cancer patients have been facing	4
2.2 Information seeking and giving behavior in clinical setting	7
2.3 Natural Language Processing Approach on clinical text	9
Methods	11
3.1 Data Source	11
3.2 Data Analysis	12
Results	17
4.1 Information seeking behavior	17
4.1.1 Information seeking behavior by Method 1	17
4.1.2 Information seeking behavior by Method 2	18
4.2 Information giving behavior	19
4.2.1 Information giving behavior by Method 1	19
4.2.2 Information giving behavior by Method 2	20
4.3 The comparison between Caucasian and African America group	21
4.3.1 The comparison between Caucasian and African America group by Method 1	21
4.3.2 The comparison between Caucasian and African America population by Method 2	22
4.4 Topic Analysis	24
4.4.1 Topic analysis of information seeking behavior dominated by patients in TD group (Method 1)	25
4.4.2 Topic analysis of information giving behavior from Caucasian vs African American patients (Method 2)	26
4.5 Limitations	30
Discussion & Conclusion	28
Bibliography	32

Introduction

Prostate cancer has become the most frequently diagnosed non-skin men cancer in U.S.

As the latest statistics show that there are 181,000 new cases of prostate cancer diagnosed in 2016 (Song et al., 2017). Every treatment option of prostate cancer will unlikely avoid complications and side effects. Generally, the decision on whether and how to treat prostate cancer is considered preference sensitive and largely depends on the communication between the health care provider (HCP) and patients. Having a better understanding of patient concerns and information needs can contribute to the decision-making and patient quality of life (Song et al., 2017).

During patient-centered clinical consultations, patients share their values and concerns about the potential benefits and harms of different treatment options; HCPs explain the medical condition and treatments and help patients make informed treatment decisions. Patient-HCP communication varies from one-way communication (i.e., information flow from HCPs to patients only) to two-way communication (i.e., both HCPs and patients share the communication content, responsibility, and their preferences and values) (Song et al., 2015).

The purpose of this study is to investigate the communication patterns (i.e., information seeking and information giving behavior) between prostate cancer patients, their family members and HCPs during consultation visits.

Specifically, this study is designed to answer the following research questions:

(1) Does the decision aid promote information seeking and giving behavior of roles (HCPs, patients and its family members) in consultation visits?

(2) Are there communication style differences from specific target population in the consultation visits?

This study will help improve the communication between HCPs, prostate cancer patients, their family members for informed treatment plan decision making by providing insight on their information seeking and giving patterns.

Literature Review

The literature review covers three major areas: the decision making challenges that prostate cancer patients have been facing, previous studies on information seeking and giving behavior in clinical settings, and natural language processing methods related to text script mining and analysis.

2.1 The decision making challenges that prostate cancer patients have been facing

Prostate cancer is the most frequently diagnosed cancer type for men. In the United States, with 181,000 new cases diagnosed in 2016; approximately 90% of prostate cancer patients have localized or regional disease (Song et al., 2017). Although there are many treatment options for patients to choose, currently there is no consensus regarding the efficacy of routine screening, nor has one treatment modality been demonstrated as the best choice. Patients and their family members (i.e., mostly are their spouse) are asked to choose from several options: radical prostatectomy, radiation therapy, or the “watch and wait” approach (O’Rourke, 1999).

Since there is no strong quality evidence and little professional consensus on one specific treatment over another, it is impossible to create clinical guidelines regarding treatment. A previous study (Patel, Mirsasdraee, & Emberton, 2003) revealed that the process of choosing treatment options was based on input from a large number of factors; there was little evidence regarding which type of input exerts the greatest influence on patients. The input with most bias may have greatest influence on patients’ decision making. In general, there is a paucity of information on how patients with prostate cancer have used

different types of input in the treatment decision making process. Based on Patel et al.'s study (2003), factors that influence treatment decisions made by men with prostate cancer include patient factors; family, friends, acquaintances and others, general practitioner; hospital doctors and nurse specialists.

However, Aning, Wassersug, & Goldenberg's study (2012) found that patients repeatedly saw their health care providers as their most important source of information, and they relied on their HCPs to inform them of management options. The HCPs still appear to have the most direct influence on patient decision making. Nevertheless, how long this status will take is uncertain. Since once being aware of their diagnosis, most patients prefer to take an active or collaborative role with their physician in decision-making. If the patients can be well guided through training books or other interventions (decision aids), they can be more active in the decision making process.

For example, Patel et al. (2003) found men with newly diagnosed prostate cancer who were assisted in obtaining information on the diagnosis were significantly more active in medical decision making and reported lower levels of anxiety. However, the degree of involvement varied from patient to patient. The observation from Cassileth et al.'s study (1989) showed that prostate cancer patients were satisfied with their treatment choices over time and suggested that patients can indeed play an active role in making intelligent decisions about their own care and treatment options and that they remain comfortable and satisfied with their choices.

Besides patients and HCPs, family members also influence medical decision making. The research conducted by O'Rourke (1999) found that decisions regarding treatment were based on a combination of the shared and individual backgrounds, biases and coping

styles of patients and their spouses. However, family members are not always consistent in the decision making. In addition, besides HCPs' professional advice, they also used informal networks from other family members and lay literature as preferred information sources in the United States.

What's more, the culture issues also should be considered into treatment decision making, such as patient's personal education background, age, race. To identify these factors can help increase the attendance in the consultation visits.

Besides the communication itself, researchers found that a good decision aid which can meet the information need of prostate cancer patients and their family members. Based on patients' preference study, Aning et al.(2012) found that the information needs of prostate cancer patients were not well met and decision-making aids were a positive treatment adjunct both to convey information and to allow patients to explore their own beliefs and values during the decision-making process. The results suggested that decision-making aids better prepared patients for involvement in treatment decisions. Lin, Aaronson, Knight, Carroll, & Dudley (2009) also found that the use of decision aids can help patients gain more relevant knowledge and more actively involved in decision-making, and decrease levels of anxiety and distress. In general, Lin et al. (2009) thought that consistent use of decision aids may help fill the information void and allow for more efficient and more effective discussions between health care providers and patients; the decision aids improved both subjective and objective knowledge at the same time. In addition, it is important that decision aids also provide guidance and assistance to patients with regard to clarifying their values and preferences for treatment.

In conclusion, unfortunately, there is no one best treatment option for prostate cancer patients. The patients have to decide the treatment option for themselves. In the consultation visits for treatment options the communications between patient, doctors, and family members play a very important role in treatment decision making.

2.2 Information seeking and giving behavior in clinical setting

The reviewed literatures discussed the information needs of patients and their family members, health-related information seeking behavior, and specifically, the HCPs' information giving behavior.

2.2.1 information needs of patients and family members

Prostate cancer patients and their family members have enormous and a variety of information needs; however, in many occasions, their information needs are not met although they're satisfied with the information already acquired. Echlin & Rees (2002) first described the distinct information needs and information-seeking behaviors that men with prostate cancer have throughout their cancer journeys. Then, they stated that although there is considerable variation in the amount and type of information that men required, the majority of men with prostate cancer were satisfied with the information they receive. However, Echlin & Rees (2002) showed that thirty percent of men and 25% of family members/friends stated that they did not receive timely and appropriate medical information during the investigation of their prostate cancer. Lin et al. (2009) also found that although men with prostate cancer have enormous needs for information, these needs are often unmet. Nevertheless, Wallston & Maides (1976) believed that the more information a person has about a threatening condition, the more likelihood this person will take positive steps to ameliorate the condition.

2.2.2 Information seeking behavior

According to Wallston & Maides (1976), the definition of information seeking behavior is a chain of behaviors which ultimately might lead to positive or negative consequences in the process of consultation visits. Although relevant information is generally available from a wide variety of sources, individuals differ greatly in the extent to which they seek and subsequently make use of such input.

Cancer patients always have an intention to talk about their own condition and the extent of information seeking depends on multiple factors. Borgers et al.'s study (1993) found a large number of cancer patients indicated an intention to discuss one or more aspects of their illness and treatment with the HCPs. The factors that might influence patients to seek information from the specialist at the outpatient clinic is uncertainty, fear and dissatisfaction with information received. In general, the information-seeking behavior of cancer outpatients appears to be influenced by several factors, including patients' needs, values and beliefs; unexpected situations; patients' skills; and specialists' and companions' behavior.

2.2.3 Information giving behavior

For the information giving behavior, the literature focused more on health care providers' information giving behavior. A meta-analysis(Hack et al., 2012) summarized that there are five categories in physician-patients' communication: information giving, question asking, partnership building, rapport building and emotional support. Both Hack et al. (2012) and Borgers et al. (1993) believed that physician scored too high on information giving behavior that it limited patients-initiated information giving discussion.

2.3 Natural Language Processing Approach on clinical text

Natural language processing (NLP) and information extraction (IE) are commonly applied within health related text in terms of processing large quantities of unstructured text and returning structured information (Jackson et al., 2017). A large number of tools and frameworks exist for general purposes of information extraction from clinical dictionaries, such as cTAKES, NOBLE and MedLee.

cTakes is an open-source natural language processing system for information extraction from electronic health record clinical free-text. It can process clinical notes, identify types of clinical named entities such as drugs, diseases/disorders, signs/symptoms, anatomical sites and procedures. In Hong et al.'s (2016) and Weng's (2017) studies, cTakes was used to extract key features from datasets. Similar to cTakes, MedLEE has been used to extract, structure, and encode clinical information in textual patient reports so that the data can be used by subsequent automated processes.

NOBLE is a text mining tool and API that automatically codes free text with concepts from controlled terminology. Its algorithm is similar to IndexFinder (Zou, Chu, Morioka, Leazer, & Kangarloo, 2003) or ConceptMapper (Tanenblatt, Coden, & Sominsky, 2009), but unlike these systems, NOBLE Coder uses NoSQL for data storage which enables it to code with huge terminologies while keeping its in-memory footprint small. Tseytlin, Mitchell, & Legowski (2016) used NOBLE Coder to implement a general algorithm for matching terms to concepts from an arbitrary vocabulary set. Compared to Ctakes, NOBLE Coder performed better on speed and accuracy.

Among professional text mining tools, Python and NLTK package are the most popular ones. Ruano (2018) used both R and Python to perform text mining, machine learning, deep learning analyses, and data visualization.

In addition, for specific bibliometric text mining, VOSviewer (Yu & Hayes, 2018) is capable of generating topic network mapping and conducting visual analytics.

Methods

3.1 Data Source

My Master Paper advisors provided me with the transcripts of the consultation visits involving health care providers, prostate cancer patients, and their family members, which I used as the original dataset for this study. One of my master paper advisors (i.e., Dr. Lixin Song) used this dataset in her previous studies on the decision aid for prostate cancer patients (Song et al., 2015, 2017). This study reused the same dataset to address the information seeking and giving issues between patients, doctors, and family members. The original dataset was audio-recorded and de-identified. To gain access to and handle the original dataset, I was added to the IRB application for this study (IRB#14-0750) and obtained a certification of the Collaborative Institutional Training Initiative (CITI) program at the University of North Carolina at Chapel Hill.

In the original dataset (Song et al., 2017), all the participants were from a randomized clinical trial (RCT) which was for testing the effects of a decision aid intervention in the southeast of USA. In the RCT, patients were randomly divided into 3 groups: a control group with usual care (control), intervention directed to the patient (TD), and intervention directed to the patient and family support person (TS).

The intervention tool adopted in the previous studies (Song et al., 2015, 2017) included a DVD that introduced communication strategies about how to engage in effective communication using information giving, seeking, verifying and clarifying skills, a booklet that provided patients a case-based guide to treatment issues for early stage

prostate cancer, and 4 telephone calls from a trained nurse interventionist who was responsible for answering patients' all kinds of questions. The patients and family members in the TS group received 4 telephone calls from the same nurse respectively. About the participant recruiting, patients were eligible if they (1) were newly diagnosed with localized prostate cancer; (2) were identified at least 10 days before the treatment consultation appointment; (3) had no major cognitive impairment; (4) had no prior cancer history; (5) could read and speak English; and (6) had a family member who was identified as providing primary support to the patient (PSP) and was willing to participate in the study. All intervention and control groups contained both African-American and Caucasian men with prostate cancer.

After recruiting participants and transcribing their audio-recorded consultation visits into texts, in the final dataset, there are 171 consultation visits ($N = 60$ in TD; $N = 57$ in TS; and $N = 54$ in control group) in total. Furthermore, I divided the control group into two additional subgroups by race ($N = 38$ in Caucasian and $N = 16$ in African America).

3.2 Data Analysis

The first step is to clean the dataset. I transformed the original dataset format from .doc format to .txt to facilitate data processing in Python. Then, I used Python to categorize every role (i.e., patient, family member, HCP) in each consultation visit. The processed conversation script clearly differentiated the roles into "Doctor: ...", "Patient: ...", "Family Member: ...".

The next step is to do the specific data analysis. The information seeking behavior was examined by analyzing the questions from doctors, patients and family members in the transcript. The information giving behavior was examined by analyzing statements from

different roles (i.e., patients, family members, and HCPs). The statistical counts of both the questions representing information seeking and the statement representing information giving instances help were generated to investigate the information seeking and giving behaviors of different roles (i.e., patients, doctors and family members). For text analysis purposes, the question marks in the original text were used to signal the information seeking behavior and the periods were used to signal information giving behavior. In addition, two methods were adopted to investigate the patterns of information seeking and giving behaviors.

Method 1 was adopted to reveal which role's information seeking or giving behavior dominated a consultation visit. If a role (i.e., patient, doctor, or family member) has more questions or statements than the others, this role was regarded as the dominant role during a consultation visit. Since family members' participation was relatively low in the consultation visits, the results of Method 1 automatically excluded the role of family members. In the final statistics, the data is shown as the percentage of patients who dominated the consultation in information seeking/giving and the percentage of doctors who dominated the consultation in information seeking/giving. Take TD group as an example, in the information seeking domination statistical chart, the percentage of patients who dominated the consultation in information seeking = the number of consultation visits in which patients dominated the information seeking / the total number of consultation visits in TD group; The percentage of doctors who dominated the consultation in information seeking = the number of consultation visits in which doctors dominated the information seeking / the total number of consultation visits in TD group. In the information giving domination statistical chart, the percentage of patients who

dominated the consultation in information giving = the number of consultation visits in which patients dominated the information giving / the total number of consultation visits in TD group; The percentage of doctors who dominated the consultation in information giving = the number of consultation visits in which doctors dominated the information giving / the total number of consultation visits in TD group.

Method 2 was adopted to show the distribution of information seeking and giving behaviors among different roles in each group by counting the total number of question marks and periods in each group (i.e., TD, TS, and Control group) from each role's transcript. In the final statistics, the data is shown as the percentage of information seeking/giving behavior by patients/family members/doctors. Take TD group as an example, in the information seeking behavior chart, the percentage of patients' information seeking behavior = the number of questions asked by patients / total number of questions in TD group; The percentage of family member's information seeking behavior = the number of questions asked by family members / total number of questions in TD group; The percentage of doctors' information seeking behavior = The number of question asked by doctors / total number of questions in TD group (Note: Total number of questions in TD group = The number of questions asked by patients + The number of questions asked by family members + The number of question asked by doctors).

In the information giving behavior chart, the percentage of patients' information giving behavior = the number of statements said by patients / total number of questions in TD group; The percentage of family member's information giving behavior = the number of statements said by family members / total number of questions in TD group; The percentage of doctors' information giving behavior = the number of statements said by

doctors / total number of questions in TD group. These two methods aimed at seeing if there is significantly difference in information seeking/giving behavior across the three groups.

Comparing Method 1 and 2, Method 1 is a micro-level analysis focusing on the information seeking/giving behavior demonstrated at each consultation visit while Method 2 is a macro-level analysis focusing on the overall information seeking/giving behavior in all consultation visits for each study group. Method 1 aimed at revealing the common behavior pattern case by case and extracting the domination role by binary counting. It can give a more direct way to see the effect of intervention on patients but not for family members because family members' participation was relatively low.

Method 2 tried to explore the overall behavior pattern in each group by full counting. It didn't consider the different conversational conditions and patients' personal demographics may play a role in affecting the information seeking and giving.

A Python script was developed to extract the number of questions and statements from each consultation transcript and analyze information seeking and giving patterns.

Statistical charts were generated in excel based on the counts generated by data processing in Python.

Moreover, this study conducted a topic analysis for patients' information seeking and information giving behavior by using VOSviewer (<http://www.vosviewer.com/>).

VOSviewer is a software for constructing and visualizing bibliometric networks. It also offers text mining functionality that can be used to construct and visualize co-occurrence networks of important terms extracted from a body of scientific literature. Natural

language processing techniques are available in VOSviewer for creating term co-occurrence networks based on English-language textual data. Relevant and non-relevant terms can be distinguished algorithmically. I chose to use this software to calculate the frequency and co-occurrence of the terms in the transcript and produce visualization analytics. The visualization showed the most frequently occurred terms and clustered these terms by co-occurrences. The purpose of topic analysis is to see the common concerns and topics rising from patients concerns and needs.

Results

4.1 Information seeking behavior

4.1.1 Information seeking behavior by Method 1

Regarding information seeking behavior investigated by Method 1, the result (Figure 1) shows that in TD group, there are 33% consultation visits in which the patients dominated the information seeking; there are 14% consultation visits in which the patients dominated the information seeking behavior in TS group; there are 25% consultation visits in which the patients dominated the information seeking behavior in the control group. Compared to both TS and control group, there are higher percentage of patients who engaged in information seeking in the TD group.

From the doctors' information seeking, there are 86% consultation visit in which the doctors dominated the information seeking in TS group, which is higher than 67% in TD group and 75% in the control group.

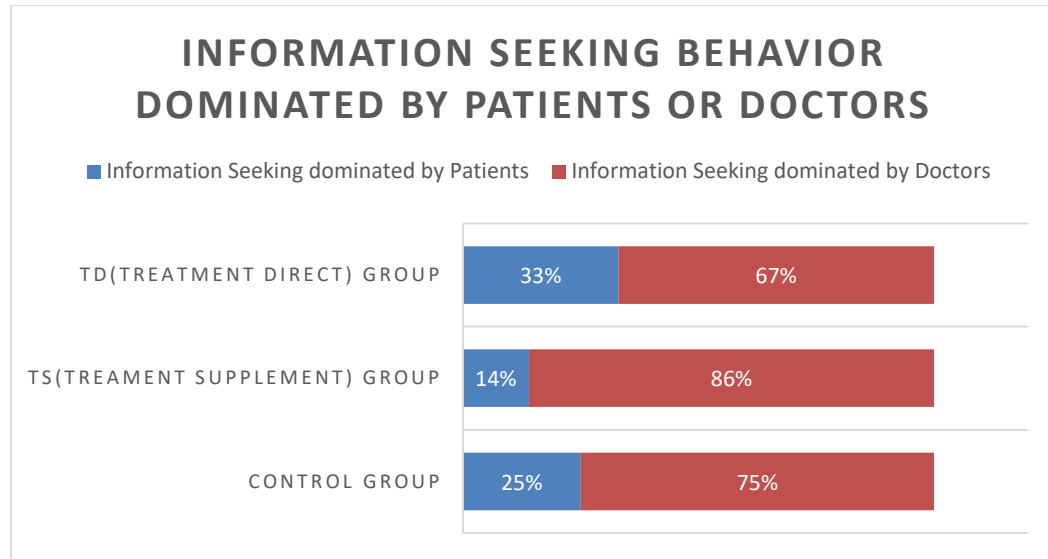


Figure 1 Information seeking behavior domination by patients or doctors

4.1.2 Information seeking behavior by Method 2

Regarding the distribution of information seeking behavior investigated by Method 2, the result (Figure 2) shows that the percentages of patients' and doctors' information seeking are almost the same across three groups, which means the total number of asked questions in each group is about the same. However, there are higher percentage of family members who engaged in information seeking in TD group (8%) and TS group (15%) than in the control group.

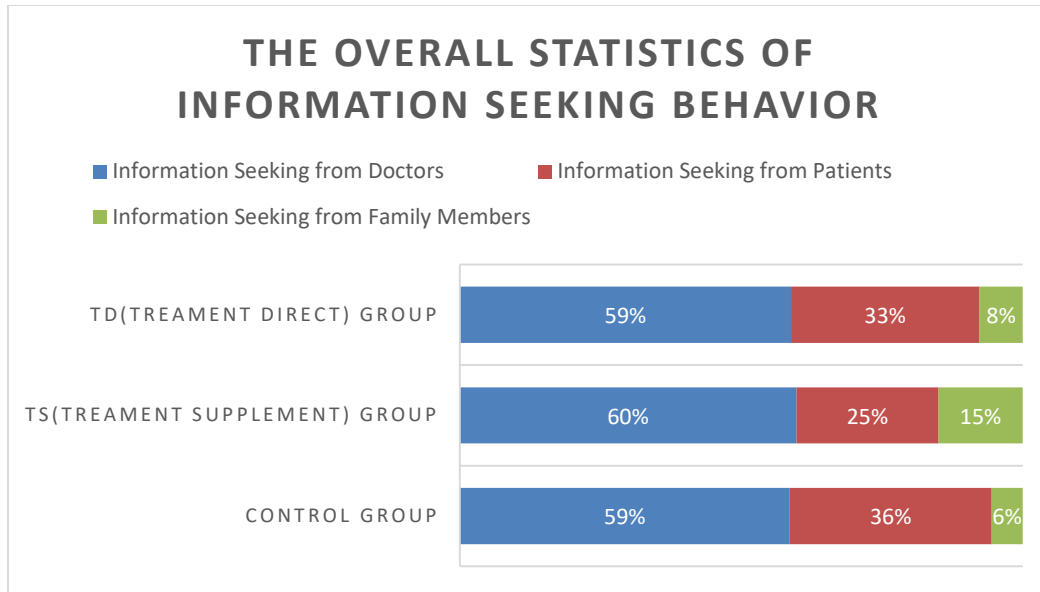


Figure 2 The overall statistics of information seeking behavior

4.2 Information giving behavior

4.2.1 Information giving behavior by Method 1

Regarding information giving behavior examined by Method 1, there were higher percentage (i.e., 10%) of patients who engaged in information giving behavior in the TD group than 5% in the control group (Figure 3). Due the binary counting in Method 1, no patients dominated information giving behavior during a consultation visit in TS group. From the doctors' information giving, doctors completely dominated the consultation visits in TS group. In the TD and control group, there are respectively 90% and 95% of consultation visits in which doctors dominated information giving.

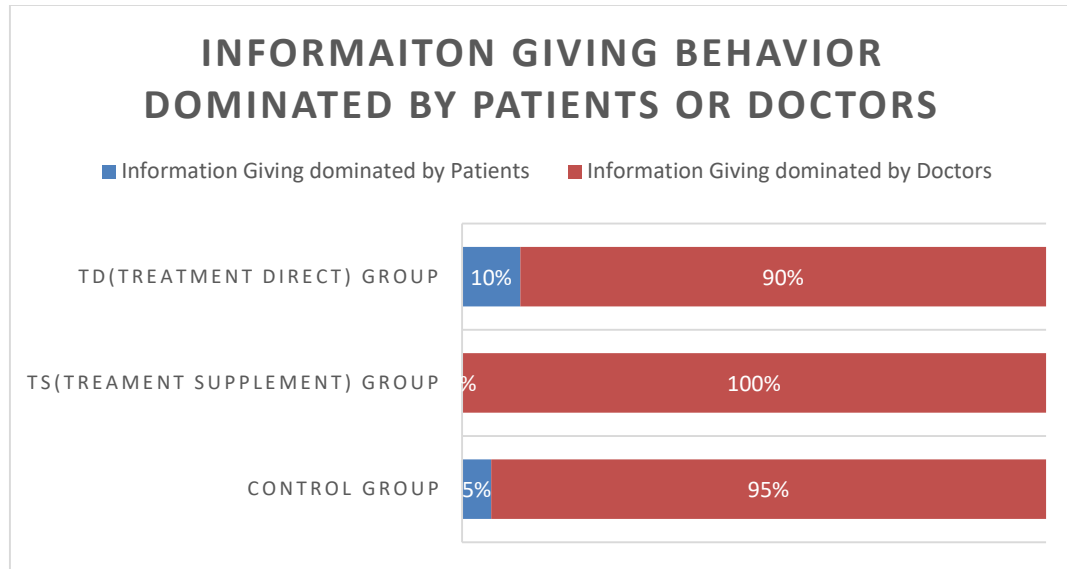


Figure 3 Information giving behavior domination by patients or doctors

4.2.2 Information giving behavior by Method 2

Regarding the information giving behavior examined by Method 2, the percentages of overall information giving by both patients and HCPs across three groups were about the same in terms of the total number of statement sentences. However, family members demonstrated slightly higher information providing percentage (8%) in TS group compared to TD (5%) and the control group (5%).

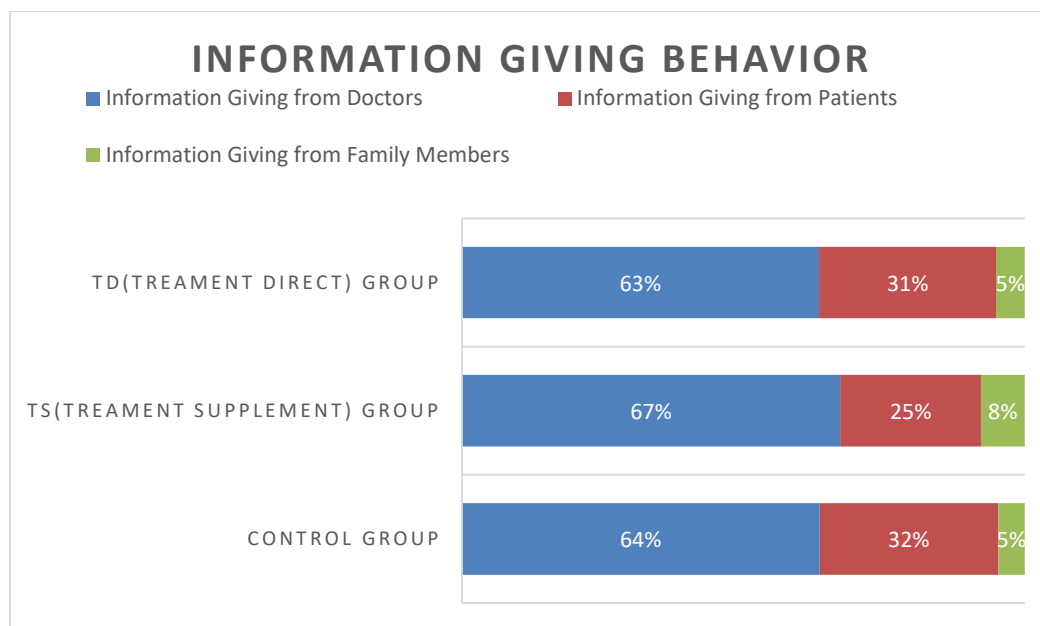


Figure 4 The overall statistics of information giving behavior

4.3 The comparison between Caucasian and African America group

The control group was selected to compare the information seeking/providing behavior difference in different races (Caucasian and African America) because the control group just followed the protocol of routine medical practice.

4.3.1 The comparison between Caucasian and African America group by Method 1

Using method 1, there are higher percentages of patients who engaged in information seeking (31%) and information giving (8%) behavior in the Caucasian group than in the African America group (Figure 5 & 6). In addition, African American patients did not dominate the information giving in any of the consultation visits.

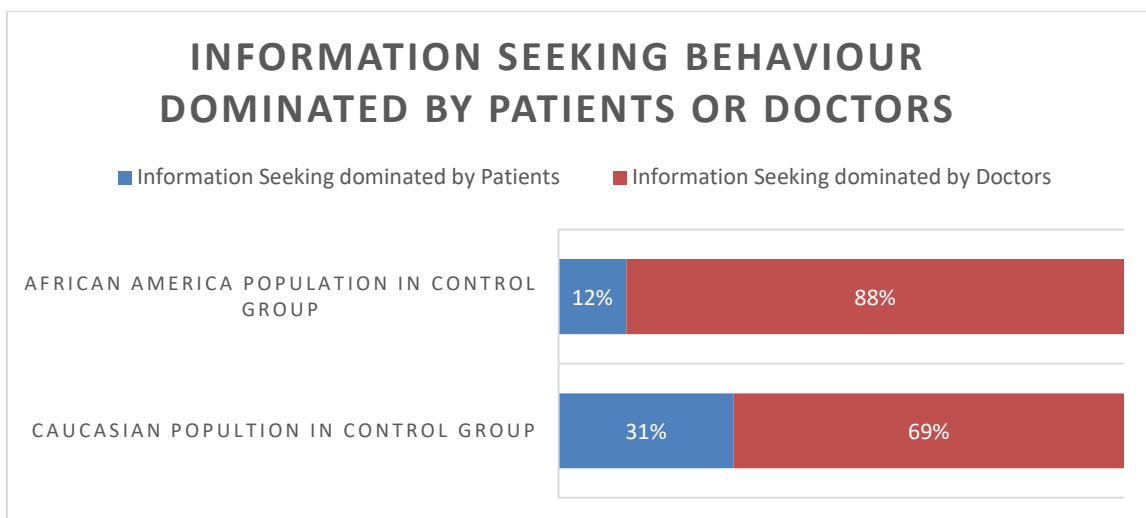


Figure 5 Information seeking behavior domination by patients or doctors in difference race

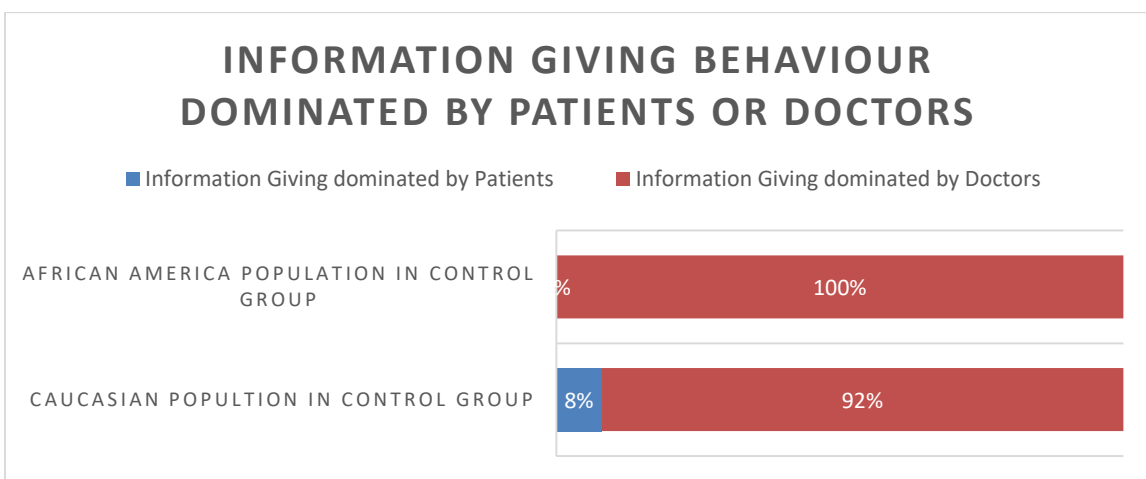


Figure 6 Information giving behavior domination by patients or doctors in different race

4.3.2 The comparison between Caucasian and African America population by Method 2

Using method 2, the results show that, compared to the African America patients (30%), there are higher percentage (38%) of Caucasian patients engaging in information seeking (Figure 7 & 8). Regarding information giving behavior, higher percentage (33%) of Caucasian patients engaged in information giving than the African America patients (28%) (Figure 9 & 10).

Therefore, for both information seeking and giving, Caucasian patients demonstrated higher percentage of engagement than the African America patients.

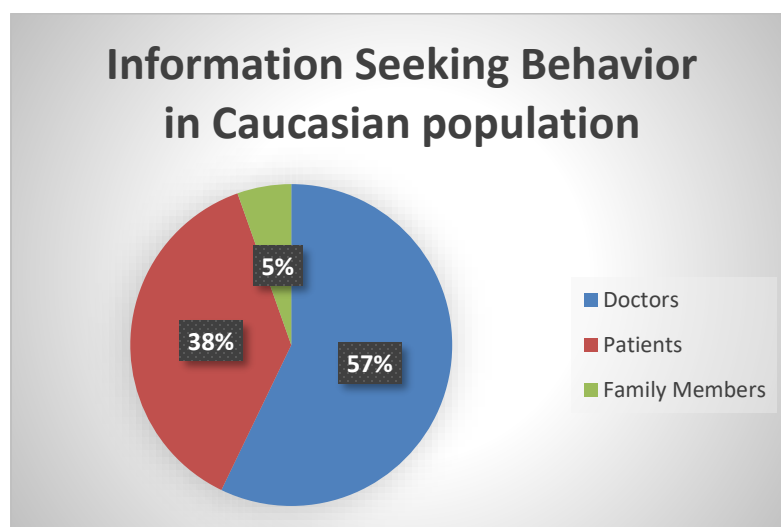


Figure 7 Information seeking behavior proportion in Caucasian population

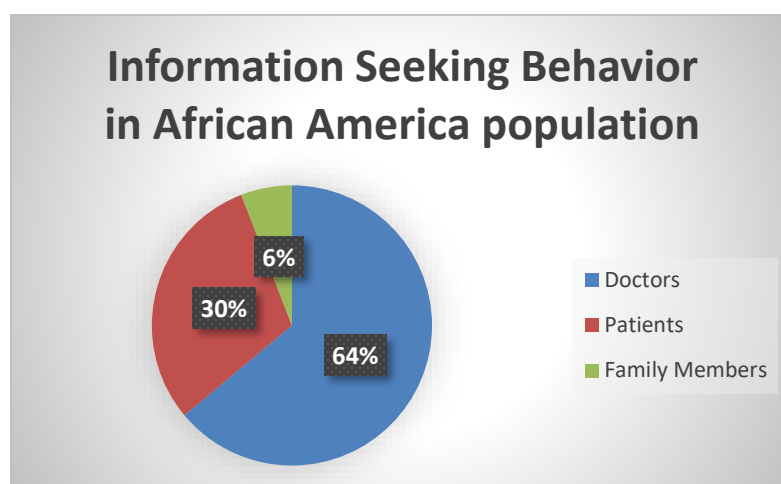


Figure 8 Information seeking behavior proportion in African America population

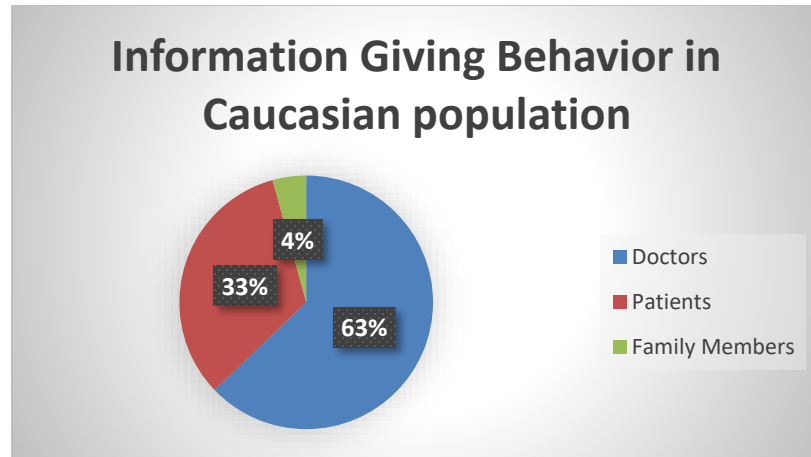


Figure 9 Information giving behavior in Caucasian population

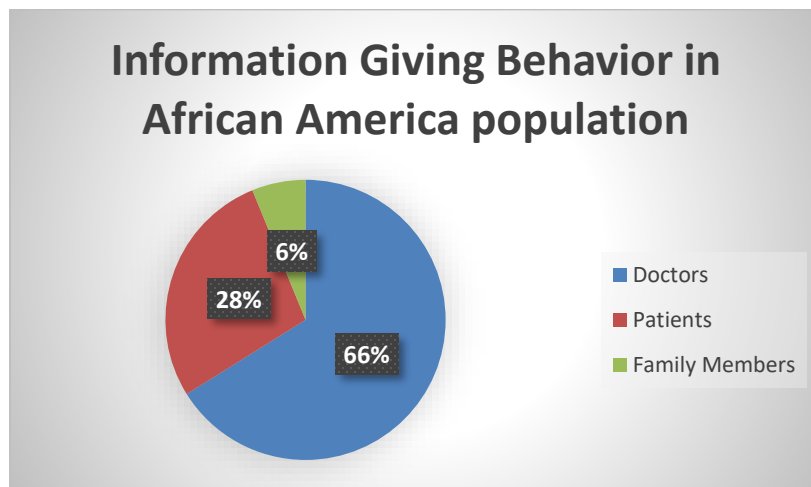


Figure 10 Information giving behavior in African America population

4.4 Topic Analysis

Since the results show that generally more patients in TD group engaged in information seeking during each consultation visit than TS and control group (Figure 1), The questions asked by patients in TD group were selected for further topic analysis. In addition, the information giving by African American patients and Caucasian patients during consultations were analyzed and compared as well.

4.4.1 Topic analysis of information seeking behavior dominated by patients in TD group (Method 1)

Overall, the topics of patients' information seeking in TD group spread to five main areas: treatment options, risks & benefits, treatment preference, diagnosis and others. In each area, there are terms that occurred in high frequencies. For example, in the area of treatment options, patients were most concern with seeds implant treatment, beam radiation treatment and medicine treatment. In the area of risks & benefits, the patients concerned over the side effect of the specific treatment options and possible pain. The results show that the patients who dominated the conversation in a consultation visit are more likely to ask specific treatment options and its side effects (Figure 11).

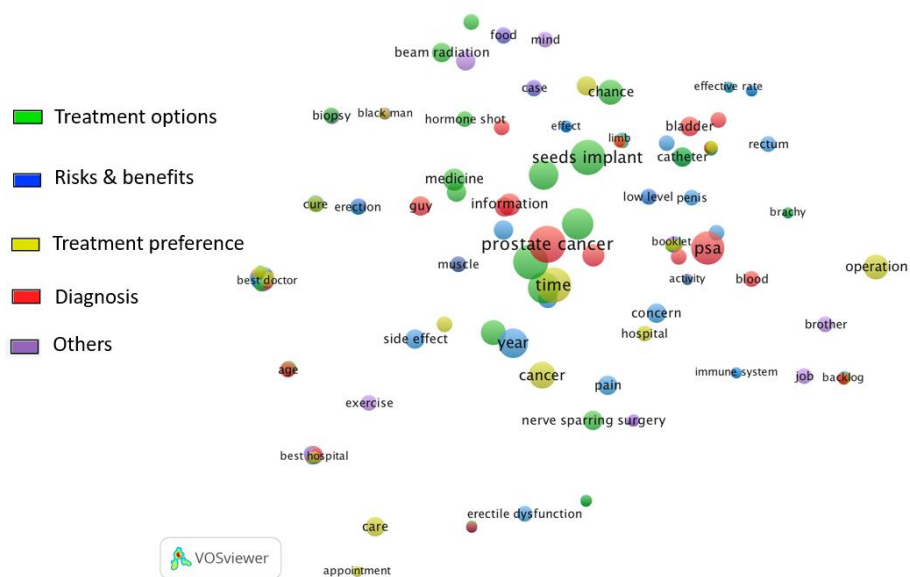


Figure 11 Topic analysis from patients who dominated information seeking behavior in TD group

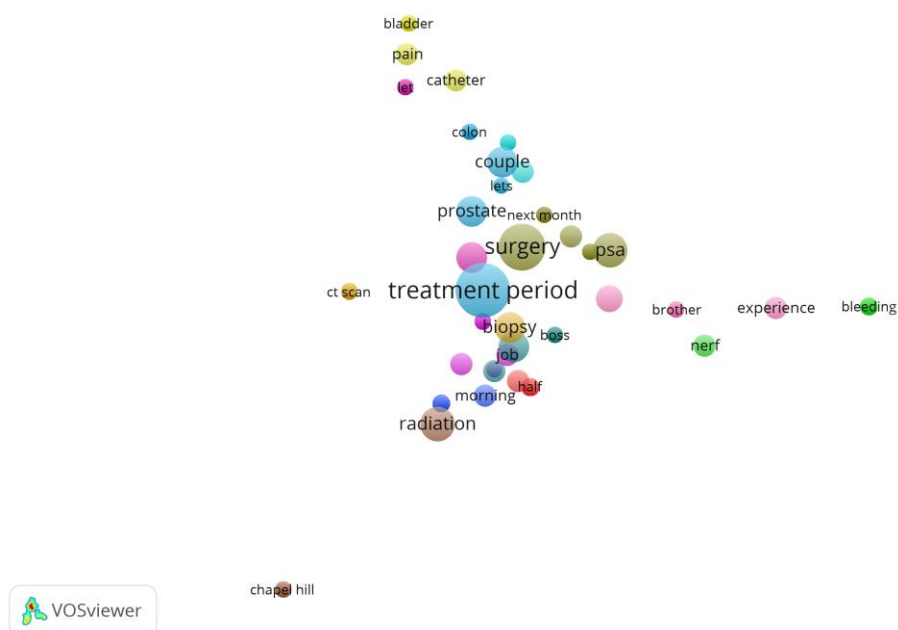


Figure 13 Topic analysis from African America population information giving behavior

Discussion & Conclusion

This study investigated the information seeking and giving behavior of patients, HCPs and their family members in three groups by two methods. The results from Method 1 investigation clearly show that the decision aid intervention did play a very important role in encouraging the information seeking behavior of patients during the consultation visits, which is consistent with a previous study (Song et al., 2017). One justification provided by the previous researcher may also applied to this study (Song et al., 2017, p. 6),

“it is likely that prior to their consultation visits, patients and their family members, when both were involved in the intervention, received more information from the multi-component intervention about prostate cancer that was pertinent to the patient’s condition. They may also have discussed the related issues with the intervention nurse and between themselves, especially the psychosocial issues related to prostate cancer. These information and discussions ahead of the consultation visit, in turn, may have resulted in a reduction in their subsequent requests for additional information from physicians than what had been provided during consultation.”

However, the results produced from adopting Method 2 show that the patients’ information seeking and giving behavior did not have significantly difference across three groups. One possible explanation is that overall, patients’ personality, characteristics, education levels, and communication styles, etc. are the key factors on their information seeking and giving behavior, which is independent from decision aid intervention.

Moreover, this study also found a relatively higher percentage of patients and family members in the control group demonstrated the information seeking and giving behavior compared with TD and TS group. One explanation was “in the original random clinical

trial, patients in the control group received information about how to stay healthy during treatment for prostate cancer, which may have increased patients' and family members' concerns about the treatment procedures and thus may have led them to ask more related questions for physicians to clarify. Maybe because of lacking sufficient information about prostate cancer and its treatments have brought these patients and their family members to ask more for the cancer itself" (Song, 2017, p. 7). The findings from this study support that the adoption of the communication training and professional aid contribute to the patients' participation in the process of treatment.

This study revealed the commonly concerned topics discussed by patients in the TD group include treatment options, risk & benefits, treatment preference, diagnoses and others. This disclosure can help health professionals improve the design of decision aid intervention in terms of creating targeted education contents for the prostate cancer patients. For example, the booklet or DVD can highlight seeds implant and beam radiation which appeared to be frequently discussed treatment options during consultation visits. In addition, the identified key topics in risks & benefits of treatment options can prompt HCPs to prepare their information giving during a consultation visit.

Furthermore, this study found the Caucasian population are more active in information seeking and information giving than the African America population. Particularly, the topic analysis showed Caucasian patients tended to use more professional terms about drug use effects and treatment option in their information giving (i.e., testosterone) and they were more knowledgeable about the available treatment options and process for prostate cancer. On the contrary, the African America patients used more laymen terms in their discussion. Regarding the topics of information seeking and giving, African

America patients care more about the treatment cost, the possible effect on their job, and the experience after treatment. Based on the results, this study would suggest the decision aid and communication style shall be customized to target different race population and education levels.

In conclusion, this quantitative study disclosed the latent information seeking and giving patterns of both patients and HCPs through a natural language processing approach. The findings not only confirm some of the results from previous studies (Song, et al., 2015, 2017), but also revealed new aspects in patient-HCP communication. The results will help healthcare professionals improve the decision aid intervention, facilitate the patient-HCP communication, and deliver more effective and personalized consultation services.

Limitations

This study has several limitations. First, since the original dataset was audio recorded and transcribed to a text format (i.e., word document), some content was missing during the transition process. In addition, there is a risk that the data cleaning process may cut the dialogs out of the context and cause errors in data analysis.

Second, the sample size of consultation visits is small (only 171 consultation visits). the results and conclusion from this study may not be generalizable.

Third, this dataset is created between 2004 and 2008. The communication style may change significantly over the ten-year period. The findings may not be applicable to the current medical communication practice.

Last but not least, these two research methods have their limitations as well. Method 1 automatically excluded the family member participation for analysis because the calculation is binary counting and focus on the role domination during a visit. Method 2

took family member participation into consideration but can hardly accommodate the factors of individual demographical difference. Some patients may like asking and talking more naturally and some tend to ask or talk less no matter they receive decision aid intervention or not. The high number of questions from patients asking more is possibly neutralized with the low number of questions from patients asking less in a group, which lead to insignificant differences. Future studies are expected to explore new natural language processing methods in analyzing information seeking and giving patterns during patient consultation visits.

Bibliography

- Aning, J., Wassersug, R. J., & Goldenberg, S. L. (2012). Patient preference and the impact of decision-making aids on prostate cancer treatment choices and post-intervention regret. *Current Oncology*, 19(S3), 37–44.
<https://doi.org/10.3747/co.19.1287>
- Borgers, R., Mullen, P. D., Meertens, R., Rijken, M., Eussen, G., Plagge, I., ... Blijham, G. H. (1993). The information-seeking behavior of cancer outpatients: a description of the situation. *Patient Education and Counseling*, 22(1), 35–46.
[https://doi.org/10.1016/0738-3991\(93\)90087-D](https://doi.org/10.1016/0738-3991(93)90087-D)
- Cassileth, B. R., Soloway, M. S., Vogelzang, N. J., Schellhammer, P. S., Seidmon, E. J., Hait, H. I., & Kennealey, G. T. (1989). Patients' choice of treatment in stage D prostate cancer. *Urology*, 33(5 Suppl), 57–62. Retrieved from
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=2523612
- Echlin, K. N., & Rees, C. E. (2002). Information needs and information-seeking behaviors of men with prostate cancer and their partners: a review of the literature. *Cancer Nursing*, 25(1), 35–41. <https://doi.org/10.1097/00002820-200202000-00008>
- Hack, T. F., Ruether, J. D., Pickles, T., Bultz, B. D., Chateau, D., & Degner, L. F. (2012). Behind closed doors II: Systematic analysis of prostate cancer patients' primary treatment consultations with radiation oncologists and predictors of satisfaction with communication. *Psycho-Oncology*, 21(8), 809–817.
<https://doi.org/10.1002/pon.1984>
- Hong, N., Li, D., Yu, Y., Xiu, Q., Liu, H., & Jiang, G. (2016). A computational framework for converting textual clinical diagnostic criteria into the quality data model. *Journal of Biomedical Informatics*, 63, 11–21.
<https://doi.org/10.1016/j.jbi.2016.07.016>
- Jackson, R. G., Patel, R., Jayatilleke, N., Kolliakou, A., Ball, M., Gorrell, G., ... Stewart, R. (2017). Natural language processing to extract symptoms of severe mental illness from clinical text: The Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. *BMJ Open*, 7(1), 1–10.
<https://doi.org/10.1136/bmjopen-2016-012012>
- Lin, G. a, Aaronson, D. S., Knight, S. J., Carroll, P. R., & Dudley, R. A. (2009). Patient Decision Aids for Prostate Cancer Treatment. *CA Cancer J Clin*, 59(6), 379–390.

<https://doi.org/10.3322/caac.20039>. Available

- O'Rourke, M. E. (1999). Narrowing the options: The process of deciding on prostate cancer treatment. *Cancer Investigation*, 17(5), 349–359.
<https://doi.org/10.3109/07357909909032877>
- PATEL, H. R. H., MIRSADRAEE, S., & EMBERTON, M. (2003). The Patient's Dilemma: Prostate Cancer Treatment Choices. *The Journal of Urology*, 169(3), 828–833. <https://doi.org/http://dx.doi.org/10.1097/01.ju.0000052056.27609.8a>
- Ruano, J. (2018). Evaluating characteristics of PROSPERO records as predictors of eventual publication of non-Cochrane systematic reviews: A meta-epidemiological study protocol. *Systematic Reviews*, 7(1), 43.
- Song, L., Toles, M. P., Bai, J., Nielsen, M. E., Bailey, D. E., Sleath, B., & Mark, B. (2015). Patient Participation in Communication about Treatment Decision-Making for Localized Prostate Cancer during Consultation Visits, (November), 1419–1429.
<https://doi.org/10.4236/health.2015.711156>
- Song, L., Tyler, C., Clayton, M. F., Rodriguez-Rassi, E., Hill, L., Bai, J., ... Bailey, D. E. (2017). Patient and family communication during consultation visits: The effects of a decision aid for treatment decision-making for localized prostate cancer. *Patient Education and Counseling*, 100(2), 267–275.
<https://doi.org/10.1016/j.pec.2016.09.012>
- Tanenblatt, M., Coden, A., & Sominsky, I. (2009). The ConceptMapper Approach to Named Entity Recognition, 546–551.
- Tseytlin, E., Mitchell, K., & Legowski, E. (2016). NOBLE - Flexible concept recognition for large-scale biomedical natural language processing. *BMC Bioinformatics*, 17(1), 32.
- Visualizing scientific landscapes. (n.d.). Retrieved April 09, 2018, from <http://www.vosviewer.com/>
- Wallston, A., & Maides, S. (1976). Health-Related Information Seeking as a Function of Health-Related Locus of Control and Health Value It is a widely held tenet that the more information a person has about a particular Me-threatening condition , the greater the likelihood the person , wil. *Journal of Research in Personality*, 10(2), 215–222. Retrieved from isi:A1976BR18100009
- Weng, W.-H. (2017). Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. *BMC Medical Informatics and Decision Making*, 17(1), 155.
- Yu, F., & Hayes, B. E. (2018). Applying Data Analytics and Visualization to Assessing

the Research Impact of the Cancer Cell Biology (CCB) Program at the University of North Carolina at Chapel Hill. *Journal of eScience Librarianship*, 7(71).
<https://doi.org/10.7191/jeslib.2018.1123>

Zou, Q., Chu, W. W., Morioka, C., Leazer, G. H., & Kangarloo, H. (2003). IndexFinder : A Method of Extracting Key Concepts from Clinical Texts for Indexing, 763–767.