What Computers Can Teach Us About Doctor-Patient Communication: Leveraging Gender Differences in Cancer Care

Mohammad Rafayet Ali[1], Taylan Sen[1],
Viet-Duy Nguyen[1], M Ehsan Hoque[1],
Ronald M Epstein[2]
[1]*Department of Computer Science*
[2]*Family Medicine, Psychiatry, Oncology and Medicine*
[1,2]*University of Rochester*
[1,2]Rochester, New York, United States.

Reza Rawassizadeh[1], Paul R. Duberstein[2]
[1] *Department of Computer Science*
[2]*Health Behavior, Society and Policy*
[1]*Metropolitan College, Boston University*
[2]*Rutgers School of Public Health*
[1]Boston, Massachusetts, [2]Piscataway, New Jersey.
United States.

***Abstract*—Advanced cancer patients sometimes spend their final days in unnecessary distress while receiving aggressive cancer treatment that is unlikely to work. Part of this problem stems from patients having incorrect understanding of their prognosis. Although studies have identified that effective doctor-patient communication is associated with better patient outcomes, most cancer patients misunderstand their prognosis. We applied computational language analysis tools (word category and language sentiment) to identify gender-specific communication characteristics associated with improved patient prognosis understanding. Analysis of 382 conversations between oncologists and patients identified that for female doctors, discussing feelings, using positive sentiment language, and speaking in shorter turns were strongly associated with better patient prognosis understanding. For male doctors, allowing patients to speak more, discussing the future, and not focusing heavily on religion or death were important. Synchrony between the doctors and patients usage of positive sentiment language was shown to be relevant only for female doctors.**

***Index Terms*—Cancer care, Patient-centered communication, Sentiment**

I. Introduction

In our lifetimes, 1/3 of us are expected to get cancer [1]. In advanced cancer, where treatment may extend or improve quality of life but cannot cure, shared decision-making is of paramount importance [2]. Research shows that effective doctor-patient communication is correlated with better patient health outcomes [3], [4]. Patients and their families face complex treatment choices regarding aggressiveness of care, effectiveness, side effects, and expected levels of success. For a patient to select the appropriate treatment, the patient must understand his/her prognosis. Studies confirm that although patients desire choice of treatment, they often have an incomplete understanding of their prognosis [5]–[7]. Despite promising communication skills training methods for physicians [8]–[11], success has been limited [12].

This research was supported in part by grant W911NF-15-1-0542 with the US Defense Advanced Research ProjectsAgency (DARPA) and the Army Research Office (ARO) anda Google Faculty Research Award.

Research in health-care communication [13]–[15] has identified differences in communication between male and female physicians. Yet, customizing physician communication skills training based on the physician's gender has not been a primary focus of research. This inspired us to objectively analyze gender differences in doctor-patient communication features. Specifically, we used computational language analysis tools on transcripts between cancer patients and their doctors to identify which communication features are more effective for male and female physicians. The data analyzed includes transcripts of conversations between 38 oncologists (M=25, F=13) and 382 of their patients with advanced cancer as part of a large intervention study [16]. The data also included a measure of each patient's prognosis understanding and patient ratings of the physician. Word category and language sentiment analyses were applied to the office visit transcripts. In addition, we compared the weights assigned by linear regression models. The degree of synchronicity between the patient and physician's positive sentiment was also evaluated. Our comparative analysis involves statistically differentiating male and female physicians' communication features.

In summary, we identified language features correlated with better patient prognosis understanding and higher patient ratings separately for male and female doctors. The features which correlated with better patient prognosis understanding include: i) increased use of positive sentiment by female doctors, ii) increased patient speaking with male doctors, and iii) less focus on religion and death by male doctors. The features associated with better patient ratings include: i) female doctors who kept conversations long, but used fewer words per turn, ii) female doctors who used a higher frequency of words related to feelings, iii) patients who had long conversations with male doctors, and iv) male doctors who discussed the future. In addition, synchronicity of positive sentiment over the course of the conversation showed synchronicity level was uncorrelated with patient prognosis understanding and patient ratings of male doctors. However, for female doctors, a high level of

978-1-7281-3888-6/19/$31.00 ©2019 IEEE

sentiment synchronicity with the patient correlated with both higher patient ratings of their doctor's communication skills and, surprisingly, lower patient prognosis understanding.

II. Related Work

Gender differences is a subject of several behavioral studies such as social interaction [17], emotion recognition [18] and facial information analysis [19]. In health-care, gender differences have received substantial attention in communication research. Difference in treatment choices, quality of life, patient satisfaction, and emotional support were some of the major areas where gender differences have been demonstrated. Meeuwesen et al. [20] analyzed gender differences between patients and doctors in verbal communication in a study of 85 transcripts of medical interviews using a speech act coding system. Their results show male doctors are more likely to give patients advice and interpretations than female doctors. In an analysis of 537 audio recordings of primary care visits, Roter et al. [21] found that female physicians had longer conversations with patients, asked more questions, and used more positive statements than the male physicians. Similar results were found in subsequent reviews by Rotel et al. [22], [23]. In two separate studies, Hall et al. [24] analyzed the verbal and nonverbal behaviors in comparison to patient satisfaction in 621 interactions of patients with their doctors. They identified that the male patients who interacted with young female physicians were most dissatisfied. Gender differences have been demonstrated in cancer care as well. Several empirical studies [13]–[15], found that female physicians more likely to provide preventive counseling and gender-specific cancer screening regardless patients gender. Seale et al. [25] revealed that male patients focused on information regarding treatment and medical procedures. Female patients, however, focused on emotional support and concerns with the impact of illness.

Patients often overestimate their life expectancy and have overly optimistic perceptions of their prognosis. A study conducted by Garmling et al. [26] on 236 patients and 161 patient-oncologist survival ranking show that non-white patients have higher discordance rate than white patients, in the US. Temel et al. [27] conducted a study with 151 newly diagnosed lung carcinoma patients and showed that patients had an inaccurate assessment of their prognosis, which may have led to aggressive chemotherapy usage towards their end-of-life. Weeks et al. [7] studied 1193 patients diagnosed with metastatic (stage IV) lung or colorectal cancer. Their results presents that 69% of patients with lung cancer and 81% of those with colorectal cancer did not understand that chemotherapy was not curative.

The field of human computer interaction has touched on improving communication between pediatric patients and clinicians [28]. Zhang et al. [29] investigated computer methods to help patients articulate their health information needs. Hong et al. [30] utilized examined the use of computer-mediated storyboarding to enhance communication between pediatric patients and clinicians. Ni et al. [31] researched the use of a hand-held projection device to enhance information exchange in doctor-patient communications. Sen et al. [32] used data analysis techniques to investigate the characteristics of effective doctor-patient communication.

As supported by the above research summaries, substantial evidence shows that i) gender differences exist in several areas regarding doctor-patient communication, and ii) the current state of affairs in health-care often leaves many patients misunderstanding their prognosis with detrimental effects. To our knowledge, however, gender-specific communication features that promote prognostic understanding have not been characterized. In this paper, we apply computational linguistic analysis to answer this question.

III. Data Set and Computational Tools

The data used in our analysis comes from a multi-site study on doctor-patient communication [33] involving 38 (Male=25, Female=13) oncologists and 382 (Male=172, Female=210) patients. The data set consists of 382 transcripts of the conversations between the doctors and their patients from regularly scheduled office visits. All of the patients had advanced (stage 3 or 4) cancer and had seen the doctor at least once previously. The transcripts were obtained from an audio recording of the visit and were professionally transcribed. Each visit was followed by survey questionnaires for both the patient and physician. The survey included the Health Care Communication Questionnaire (HCCQ) [34] in which patients rated physicians communication behaviors, and a prognosis forecasting questionnaire [7]. The specific questions we investigated were:

- My doctor encouraged me to ask questions.
- My doctor was willing to discuss any topic of importance to me.
- My doctor gave me the information I could understand.
- I felt understood by my doctor.
- I feel that I got enough information from my doctor.

Patients gave a response on a 10-point Likert scale, 10 being the highest. We took the sum of these question responses as a single cumulative score representing patients' ratings of their doctor's communication skill. From the prognosis forecasting questionnaire [7] we have selected two questions to measure the patients' prognosis understanding. The doctors were asked: "What do you believe are the chances that this patient will live for 2 years or more?" Similarly, the patients were asked: "What do you believe your doctor thinks are the chances that you will live for 2 years or more?" For each, a response could be given in seven different probability ranges varying from 100% to 0%.

To measure the prognosis misunderstanding we take the absolute difference between the responses to these two questions. The prognosis misunderstanding score varies from 0 to 6 (i.e., 6 being the highest level of misunderstanding). In a demographics survey, both doctors and patients gave their self-gender identity; all doctors and patients were either male or female (i.e., no one responded with the option other).

From the conversation transcripts, word category usage and language sentiment were analyzed with two computational tools. The Linguistic Inquiry and Word Count (LIWC) tool

TABLE I: List of features extracted.

Text Features	LIWC Features		
Total words	Achievement	Health	Feeling
Total unique words	Affective	Hearing	Fillers
Total words per turn	Anger	Home	Friends
Doctor words count	Anxiety	Humans	Future tense
Doctors unique words count	Biological	Insight	Swear
Doctors word per turn	Body	Leisure	Time
Doctor positive sentiment	Causation	Money	Death
Patient words count	Certainty	Religion	
Patient unique words count	Cognitive	Sadness	
Patient words per turn	Discrepancy	Sexual	
Patient positive sentiment	Family	Social	

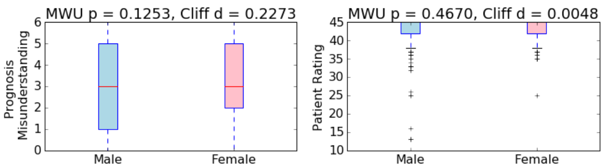


Fig. 1: Difference between male and female doctors' outcome measures (left: prognosis misunderstanding, and right: patients' rating)

[35] measures the frequency of specific word category usage in a text sample. For our analysis, we normalized the LIWC features by the total number of words spoken. Language sentiment, (i.e., the net positive emotion) associated with a sentence was evaluated using the VADER (Valence Aware Dictionary for sEntiment Reasoning) text-based sentiment analyzer [36]. In addition to LIWC and VADER, several basic statistical features of the transcripts were also calculated (see Table I).

IV. Statistical Analyses

In the first set of analyses ("Independent Analyses"), we investigated possible differences between male and female physicians with regards to language features and outcome measures (prognosis misunderstanding and patient ratings of their physicians). These analyses of language features and outcome measures were conducted independently of each other. In the second set of analyses ("Outcome Dependent Analyses"), we determined which language features were associated with better outcome measures in conversations with male physicians and separately in conversations with female physicians.

A. Independent Analyses

1) Outcome Measures : We first investigated the differences between male and female doctors prognosis misunderstanding scores and the patient ratings. We applied the Mann-Whitney U test [37], since our data does not fit to a Normal distribution. To quantify differences, Cliff's d effect size was used, appropriate for data that is not Normally distributed [38]. As shown in Fig. 1, no significant differences were found in prognosis misunderstanding and patient ratings between male and female physicians. This shows, there is no overall significant physician-gender bias with regards to the outcome.

TABLE II: Differences in features between male and female doctors' conversation. $p < 0.001$ (D=doctor, P=patient, wpt=words per turn)

Feature	Male Dr. Group Mean (std)	Female Dr. Group Mean (std)	Effect size
D words	1664 (1136)	2410 (1195)	0.42
D unique words	418 (168)	524 (164)	0.39
total words	3078 (1660)	4360 (2036)	0.38
total unique	607 (206)	739 (217)	0.35
Cognitive	0.15 (0.01)	0.16 (0.01)	0.29
Insight	0.016 (0.004)	0.018 (0.003)	0.27
P words	965 (655)	1337 (934)	0.26
D wpt	16.3 (8)	17.9 (6)	0.25
total wpt	12.3 (4)	13.2 (3)	0.23
P unique	293 (124)	348 (149)	0.23

2) Language Features: Next, we looked at the differences in the conversation features listed in Table I between conversations with male and female physicians, independent of the outcome measures. Since the language features were not normally distributed, we again used the Mann-Whitney U test and Cliff's d effect size measure. Table II lists the features, which were found to be significantly different($p < 0.05$). In order to counteract the increased likelihood of false positives when multiple hypotheses are tested, we applied the commonly used Bonferroni correction [39] in all statistical analyses. In this process, each of the p-values obtained was multiplied by the number of features, and only if the resulting p-value was less than 0.05 the results deemed significant.

As shown in Table II, the average words spoken by the doctor ("D words") was 2410 for female doctors and 1644 for male doctors ($p < 0.001$), The number of unique words spoken by the doctor ("D unique") was also higher for female doctors with an average of 524 compared to 418 spoken by male physicians. In addition, female doctors spoke with slightly more words per turn than male doctors (17.9 vs. 16.3 words per turn).

Regarding the LIWC word categories, female doctors used the cognitive, insight, and discrepancy word categories with slightly, yet significantly larger frequencies. The cognitive word category represents words pertaining to thinking such as believe, consider, and hope. The insight word category, focuses more on reflective words such as interpret and learn. The discrepancy word category involves words associated with a disparity of thoughts such as besides, rather, regret, and unnecessary.

B. Outcome Dependent Analyses

We next performed statistical comparisons between good and poor patient outcome subgroups within the male and female physician conversation group. The first set of dependent analyses investigates which language features are associated with differences between the high and low prognosis misunderstanding groups. In the second set of dependent analyses, the language features which are different between a high and low patient rating groups are compared between male and female physician conversations.

TABLE III: Differences in features between high and low misunderstanding groups for male and female doctors. $p < 0.01$

Physician gender	Feature	High PMU Group mean (std)	Low PMU Group mean (std)	Effect size
Female	Insight	0.019 (0.003)	0.018 (0.003)	0.85
	D pos	0.21 (0.05)	0.23 (0.06)	0.84
	D words	2564 (1172)	2234 (1157)	0.84
Male	P wpt	8.61 (3.3)	9.54 (3.5)	0.28
	P pos	0.30 (0.09)	0.28 (0.08)	0.26
	Death	0.0003 (0.0)	0.0002 (0.0)	0.25
	Sadness	0.0017 (0.0)	0.0020 (0.0)	0.26
	P words	879 (629)	1029 (675)	0.26

TABLE IV: Differences in features between high and low ratings groups for male and female physicians. $p < 0.01$

Physician gender	Feature	High Ratings Group mean (std)	Low Ratings Group mean (std)	Effect size
Female	Time	0.044 (0.006)	0.046 (0.006)	0.81
	Feeling	0.004 (0.002)	0.003 (0.001)	0.79
Male	Future	0.016 (0.004)	0.015 (0.004)	0.14
	Sexual	0.0004 (0.0)	0.0003 (0.0)	0.12

1) Differences between High and Low Patient Prognosis Misunderstanding Groups: The conversations with physicians were split into a high and low patient prognosis misunderstanding (PMU) groups based on the median patient prognosis misunderstanding score. For both the male and female doctor groups, this resulted in approximately equal sized groups. The language features were compared between the low and high prognosis misunderstanding groups separately for conversation data with each doctor gender.

Table III lists the language features which were different between the Low and High PMU groups for the male and female physician conversations. Female physicians who spoke with higher positive sentiment had patients with better prognosis understanding. In addition, misunderstanding was higher for female physicians who spoke more. Each of the language features pertaining patient word length (i.e., total patient words, patient words per turn) were significantly higher in the low prognosis misunderstanding group for male physicians. This suggests that when male physicians allowed their patients to speak at length and without interruption, the patients had a better understanding of their prognosis. Doctor-patient conversations with male physicians containing death-related words are more frequent in the high PMU group but sadness related words are more frequent in low PMU group. This suggests the male physicians who gave emotional support but did not focus on death were more successful.

2) Differences in High and Low Patient Ratings Groups: Table IV shows the features with significant differences between high and low rated male and female physicians. Highly rated female physicians had conversations with more words from the feeling word category and fewer words from the time word category. High rated male physicians talked with a higher frequency of words from the future and sexual word categories.

V. LINEAR REGRESSION ANALYSIS

While the previous section identified statistical differences separately one feature at a time, in this section linear regression was used to determine the relative importance of the features in predicting the outcomes. Linear regression [40] was applied to the features listed in Table I to predict the prognosis misunderstanding levels and the patient ratings separately. In addition, linear regression was separately applied on the male doctor conversation data and female doctor conversation. To minimize potential negative effects of having a large number of non-relevant features, L1 regularized variation (LASSO) of linear regression [41] was used.

A. Regression Models with the Text Features

Fig. 2a and 2b show the weights of the features assigned by the linear regression models for patient prognosis misunderstanding and patient ratings respectively. In these figures, positive weights indicate a positive relationship between the input features and the predicted the outcome.

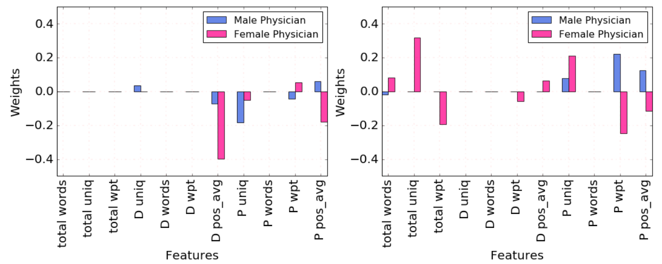
Patient Prognosis Misunderstanding. From Fig. 2a, we see that only 5 out of 11 text features have a nonzero weight. As shown, the number of doctor unique words spoken (D uniq) and patient positive sentiment (P pos_avg) both have positive weights. This indicates that an increase in the number of unique words spoken by the doctor or an increase in a patient's positive sentiment, will increase the predicted prognosis misunderstanding score. The doctors' average positive sentiment, the number of patient unique words and the number of words per turn spoken by the patient all have a negative weight, with the number of patient unique words having the largest magnitude weight.

From the female doctor regression model for patient prognosis misunderstanding in Fig. 2a, we find that only 4 out of the 11 text features had nonzero weight. Only the patient words per turn (P wpt) had a positive weight. The female doctor's positive sentiment level had the strongest magnitude weight, which was also negative. The number of patient unique words and the patient positive sentiment also had negative weights.

Patient Rating. As shown in Fig. 2b, only four text features were automatically identified as being relevant (i.e., nonzero) for male physicians. The positive weight features are the number of patient unique words, the patient word per turn, and the patient positive sentiment level. The largest magnitude of weight was patient words per turn.

For female physicians, 8 of the 11 features had nonzero weights. The positive weight features are the total number of words, the total number of unique words, the doctor's positive sentiment, and the number of patient unique words. The features with a negative weight included all of the words per turn features (i.e., total, doctor, and patient) as well as the patient positive sentiment. The largest magnitude weight was the total number of unique words. This indicate that different ways of keeping the conversation longer are associated with higher patient ratings.

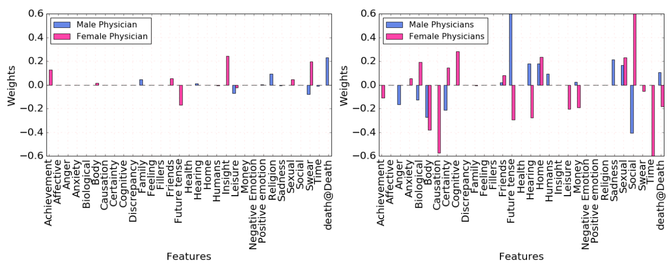
There were several notable differences in the patient rating prediction weights for male and female doctors. While the female doctor model had large negative weights for all words per turn features, the male doctor model had zeros for the words per turn features except for patient words per turn.



(a) Predicting prognosis misunderstanding

(b) Predicting patient rating

Fig. 2: Linear regression weights (D=Doctor, P=patient, pos =positive sentiment, wpt = words per turn)



(a) Predicting prognosis misunderstanding

(b) Predicting patient rating

Fig. 3: Linear regression LIWC feature weights

While the female doctor model had large positive weights for several word count features, word count features in the male doctor model were either zero or small. In addition, the patient positive sentiment was inverted between the male and female patient rating prediction models.

B. Regression Models with the LIWC Features

Next, we applied the linear regression models using the Linguistic Inquiry and Word Count (LIWC) features. Fig. 3a and 3b show the linear regression weights of the LIWC features for predicting the prognosis misunderstanding and the patients rating of the doctors respectively.

Prognosis Misunderstanding. For male doctors, the family, religious and death related words had positive weights, predicting greater prognosis misunderstanding, but similar patterns were not observed for female doctors (Fig. 3a). In contrast, in conversations with female doctors use of achievement, insight, and swear related categories were associated with greater prognosis misunderstanding.

Patient Rating. For both conversations with male and female doctors, the home, social, and sexual word category had positive weights. The future tense category had a positive weight for male doctors but negative weight for female doctors. This shows that the same recommendation for the doctors may not be appropriate.

VI. Sentiment Synchronicity Investigation

To compute the synchronicity over time, we divided each conversation into ten segments (each segment representing 10% of the conversation) and calculated the cosine similarity [42] between the doctor and patients' positive sentiment of over the sequence of segments. This provides us with a similarity measure which is invariant of the overall magnitude differences in sentiment. Overall, female doctors were significantly ($p < 0.05$) more in sync with their patients sentiment (m = 0.674, std = 0.14) compared to male doctors (mean = 0.610, std = 0.18).

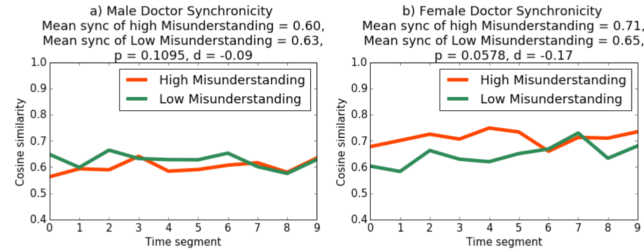


Fig. 4: Sentiment Synchronicity between doctor and patient of high and low patient prognosis misunderstanding groups.

A. Sentiment Synchronicity in High and Low Prognosis Misunderstanding

To investigate the relationship between the synchronicity and the prognosis misunderstanding, we divided the conversations into high and low PMU based on the median value. The difference in synchronicity was not significant ($p = 0.36$). In the next analysis, we divided the groups based on physician gender and then looked into the synchronicity of high and low PMU groups within each gender. For male doctors, the average synchronicity over ten segments was not significantly different between the high and low misunderstanding groups as shown in Fig. 4. However, for female doctors, the high misunderstanding group has a higher average synchronicity than the low misunderstanding group.

B. Sentiment Synchronicity in High and Low Patient Ratings

We divided the conversations into two groups high and low patients rating groups and performed the U test to find the difference in the synchronicity. The difference was not significant. We then divided the groups by physician gender and looked into the synchronicity of high and low rating groups within each gender. Fig. 5 shows the male and female doctors progression of synchronicity though duration of the conversation. There was no significant difference in the male doctors group. However, in female doctors group, the high patient rating group had higher synchronicity than the low rating group.

VII. Discussion

Female doctors had longer conversations but no differences in outcomes. This finding is consistent with prior reports [20], [24], [43]. In addition, we found that female doctors had longer turns than male doctors. Even through female doctors had a higher word repetition rate, because they spoke with a higher number of unique words, it is unclear the degree to which female doctors covered a similar number of topics as their

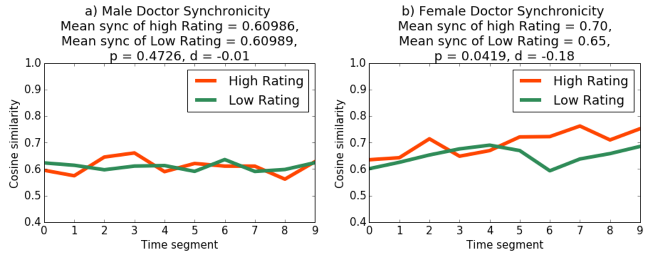


Fig. 5: Sentiment Synchronicity between doctor and patient of high and low patient rating groups.

male peers using more diverse language, or covered more topics than their male peers.

In patient-centered communication and decision-making, understanding patients emotions are considered important. Our analyses showed that in conversations with male physicians, patients' use of positive sentiment language had an inverse relationship to patients' prognosis understanding. For female physicians, the opposite effect held.

It is understandable how greater prognosis understanding could lead to negative affect, especially in advanced cancer. In a similar vein, the correlation of patient positive sentiment with poor patient prognosis understanding may signify blissful ignorance by patients, or sugarcoating by physicians. If physicians sugar-coat information rather than being more direct, it may lead to poor patient prognosis understanding. Thus, male physicians might want to be careful about keeping the emotional tone neutral, or mixed, rather than uniformly positive. A previous study on patients sentiment [32] found similar results. These behaviors may explain our findings with the male physicians, but not female physicians. The correlation of positive patient sentiment with prognosis understanding may be due to the behavior of "putting on a brave face". We don't know why such behavior may be more common in conversations with female doctors as compared to male doctors. Thus, female physicians may benefit from emphasizing the positive in order to facilitate prognostic understanding.

With male physicians, patients' use of positive sentiment language was associated with higher patient ratings. Alternatively, patient positive sentiment for female doctors was associated with lower ratings. These findings could be collectively explained if patients have a likelihood of "killing the messenger". For both the patients of male and female physicians, patient ratings were inversely correlated with prognosis understanding. Thus, patient positive sentiment might not be directly affecting patient ratings, but rather patient ratings may be affected by their level of prognosis understanding.

For male doctors, allowing the patients to speak without interruption might reduce their patients' prognosis misunderstanding and increase the patient ratings of the doctor. Whereas communication guidelines often emphasize what doctors should say about prognosis, these guidelines should be amended to emphasize that listening and having patients feel understood is equally important. In contrast, such a finding with regards to patient words per turn was not observed in conversations with female doctors.

Male doctors should take special concern when discussing death and religion. When male doctors talk about death and religion, it had an inverse relationship with patient prognosis understanding. Consistent with Terror Management Theory [44], death related conversation might increase patient anxiety, which may contribute to the avoidance of prognosis discussions when they do occur. Terror Management Theory also suggests that creating greater affiliative bonds can temper the effect of death-related anxiety, suggesting that strengthening the patient-physician relationship should precede talk about serious prognoses when possible.

We found that in conversations with male doctors, sentiment synchronicity was not associated with the outcome measures at all. In contrast for female doctors, high sentiment synchronicity was associated with mixed outcomes. While patients gave female doctors with higher synchronicity better patient ratings, this came with poor prognosis understanding, perhaps from having the illusion of understanding from greater emotional rapport. This suggests that, especially for female doctors, rapport-building should not be confused with understanding and physician training should encourage both.

The dataset was limited to conversations in the context of advanced cancers, between oncologists and patients, and in two clinical sites. Thus, extrapolating these findings to other types of medical visits with different clinical contexts and different settings (e.g., primary care visits) cannot be guaranteed to hold. Even though the number of male and female physicians differed, we used statistical tests which are valid for unbalanced data. Our analyses provide a preliminary look into gender differences in cancer communication, however, future studies should confirm our findings. Furthermore, relationship between different gender of patient and physician should be investigated. Because this was a cross-sectional study, causality cannot be ascertained, and may, in fact, be circular, with behaviors and outcomes reinforcing one another. Nonetheless, there are lessons regarding factors, previously not characterized, that may influence the difficult conversations physicians to have with their patients regarding prognosis.

VIII. Conclusion

In this paper, we looked at gender differences in cancer care communication through the lens of computational language analysis tools. We identified some key recommendations that might help improve the patient's prognosis understanding. These findings may lead to physician education that will benefit from gender-specific training. However, in observational studies, it is impossible to assign causality and there might be other factors, which contributed to the misunderstanding. We are hopeful that the value of our findings will synergistically combine with others related work.

References

[1] National Cancer Institute. Cancer Statistics, 2018.

[2] R. M. Epstein and R. L. Street. The values and value of patient-centered care, 2011.

[3] S. H. Kaplan, S. Greenfield, and J. E. Ware. Assessing the Effects of Physician-Patient Interactions on the Outcomes of Chronic Disease. *Medical Care*, 1989.

[4] M. Stewart, J. B. Brown, A. Donner, I. R. McWhinney, J. Oates, W. W. Weston, and J Jordan. The impact of patient-centered care on outcomes. Technical report, 2000.

[5] S. Parker, J. Clayton, K. Hancock, S. Walder, P. Butow, S. Carrick, D. Currow, D. Ghersi, P. Glare, R. Hagerty, and M. Tattersall. A systematic review of prognostic end-of-life communication with adults in the advanced stages of a life-limiting illness: Patient caregiver preferences for the content, style, and timing of information, 2007.

[6] A. A. Wright, B. Zhang, A. Ray, J. W. Mack, E. Trice, T. Balboni, S. L. Mitchell, V. A. Jackson, S. D. Block, P. K. Maciejewski, and H. G. Prigerson. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. *JAMA*, 2008.

[7] J. C. Weeks, P. J. Catalano, A. Cronin, M. D. Finkelman, J. W. Mack, N. L. Keating, and D. Schrag. Patients' Expectations about Effects of Chemotherapy for Advanced Cancer. *New England Journal of Medicine*, 367(17):1616–1625, 2012.

[8] V. Smothers, R. Ellaway, and C. Balasubramaniam. eViP: sharing virtual patients across Europe. *AMI Annual Symposium proceedings*, pages 1140–1140, 2008.

[9] R. C. Hubal, P. N. Kizakevich, C. I. Guinn, K. D. Merino, and S. L. West. The virtual standardized patient. In *Studies in Health Technology and Informatics*, volume 70, pages 133–138, 2000.

[10] A. Stevens, J. Hernandez, K. Johnsen, R. Dickerson, A. Raij, C. Harrison, and M. DiPietro. The use of virtual patients to teach medical students history taking and communication skills. *American Journal of Surgery*, 191(6):806–811, 2006.

[11] A. Kleinsmith, D. Rivera-Gutierrez, G. Finney, J. Cendan, and B. Lok. Understanding empathy training with virtual patients. *Computers in human behavior*, 52:151–158, 2015.

[12] W. T. Riley, P. Pilkonis, and D. Cella. Application of the National Institutes of Health Patient-reported Outcome Measurement Information System (PROMIS) to mental health research. *The journal of mental health policy and economics*, 14(4):201–208, 2011.

[13] J. T. Henderson and C. S. Weisman. Physician gender effects on preventive screening and counseling: an analysis of male and female patients' health care experiences. *Medical care*, 2001.

[14] N. Lurie, K. L. Margolis, P. G. McGovern, P. J. Mink, and J. S. Slater. Why do patients of female physicians have higher rates of breast and cervical cancer screening? *Journal of General Internal Medicine*, 1997.

[15] P. Franks and C. M. Clancy. Physician gender bias in clinical decision-making: screening for cancer in primary care. *Medical care*, 1993.

[16] R. Epstein, P. Duberstein, J. Fenton, K. Fiscella, M. Hoerger, D. Tancredi, G. Xing, R. Gramling, S. Mohile, P. Franks, P. Kaesberg, S. Plumb, and C. Cipri. Effect of a Patient-Centered Communication Intervention on Oncologist-Patient Communication, Quality of Life, and Health Care Utilization in Advanced Cancer: The VOICE Randomized Clinical Trial. *JAMA oncology*, 2017.

[17] G. Stratou, R. Hoegen, G. Lucas, and J. Gratch. Investigating gender differences in temporal dynamics during an iterated social dilemma: An automatic analysis using networks. In *Affective Computing and Intelligent Interaction*, pages 531–536. IEEE, 2017.

[18] M. Bilalpur, S. M. Kia, T. Chua, and R. Subramanian. Discovering gender differences in facial emotion recognition via implicit behavioral cues. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 119–124. IEEE, 2017.

[19] B. Xiao, P. Georgiou, B. Baucom, and S. Narayanan. Head motion modeling for human behavior analysis in dyadic interaction. *IEEE transactions on multimedia*, 17(7):1107–1119, 2015.

[20] L. Meeuwesen, C. Schaap, and C. Staak. Verbal analysis of doctor-patient communication. *Social Science and Medicine*, 1991.

[21] D. Roter, M. Lipkin, and A. Korsgaard. Sex differences in patients' and physicians' communication during primary care medical visits. *Medical care*, 1991.

[22] D. Roter, J. a Hall, and Y. Aoki. Physician gender effects in medical communication: a meta-analytic review. *JAMA*, 2002.

[23] D. Roter and J. Hall. Physician Gender and Patient-Centered Communication: A Critical Review of Empirical Research. *Annual Review of Public Health*, 2004.

[24] J. Hall, J. Irish, D. Roter, C. Ehrlich, and L. Miller. Satisfaction, gender, and communication in medical visits. *Medical Care*, 1994.

[25] C. Seale, S. Ziebland, and J. Charteris-Black. Gender, cancer experience and internet use: A comparative keyword analysis of interviews and online cancer support groups, 2006.

[26] R. Gramling, K. Fiscella, G. Xing, M. Hoerger, P. Duberstein, S. Plumb, S. Mohile, J. J. Fenton, D. J. Tancredi, R. L. Kravitz, and R. Epstein. Determinants of patient-oncologist prognostic discordance in advanced cancer. *JAMA Oncology*, 2(11):1421–1426, 2016.

[27] J. Temel, J. Greer, S. Admane, E. Gallagher, V. Jackson, T. Lynch, I. Lennes, C. Dahlin, and W. Pirl. Longitudinal perceptions of prognosis and goals of therapy in patients with metastatic non-small-cell lung cancer: results of a randomized study of early palliative care. *Journal of clinical oncology*, 2011.

[28] L. Wilcox, R. Patel, A. Back, M. Czerwinski, W. Pratt, E. Horvitz, and P. Gorman. PatientClinician Communication: The Roadmap for HCI. In *Conference on Human Factors in Computing Systems*, 2013.

[29] Y. Zhang and W. T. Fu. Designing consumer health information systems: What do user-generated questions tell us? In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011.

[30] M. Hong, U. Lakshmi, T. Olson, and L. Wilcox. Visual ODLs: Co-Designing Patient-Generated Observations of Daily Living to Support Data-Driven Conversations in Pediatric Care. *In Proc. of CHI*, 2018.

[31] T. Ni, A. Karlson, and D. Wigdor. Anatonme: Facilitating doctor-patient communication using a projection-based handheld device. In *Proceedings of CHI 2011*. ACM.

[32] T. Sen, M. R. Ali, M. E. Hoque, R. Epstein, and P. Duberstein. Modeling doctor-patient communication with affective text analysis. In *2017 7th International Conference on Affective Computing and Intelligent Interaction, ACII 2017*, 2018.

[33] M. Hoerger, R. Epstein, P. Winters, K. Fiscella, P. Duberstein, R. Gramling, P. Butow, S. Mohile, P. Kaesberg, W. Tang, and S. Plumb. Values and options in cancer care (VOICE): Study design and rationale for a patient-centered communication and decision-making intervention for physicians, patients with advanced cancer, and their caregivers. *BMC Cancer*, 2013.

[34] P. Gremigni, M. Sommaruga, and M. Peltenburg. Validation of the Health Care Communication Questionnaire (HCCQ) to measure outpatients' experience of communication with hospital staff. *Patient Education and Counseling*, 2008.

[35] J. W. Pennebaker, R. J. Booth, and M. E. Francis. Operator's Manual: Linguistic Inquiry and Word Count - LIWC2007. *Depts.Ttu.Edu*, 2007.

[36] C. J. Hutto and E. Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*, 2014.

[37] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

[38] N. Cliff. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological bulletin*, 114(3):494, 1993.

[39] R. A. Armstrong. When to use the Bonferroni correction, 2014.

[40] D. A. Freedman. Statistical Models: Theory and Practice. *Cambridge University Press*, 2009.

[41] P. Bühlmann and S. Geer. *Statistics for high-dimensional data: Methods, Theory and Applications*. 2011.

[42] C. Aggarwal. *Data Mining: The Textbook*. 2015.

[43] P. Franks and K. D. Bertakis. Physician gender, patient gender and primary care. *Journal of Women's Health*, 2003.

[44] J. Greenberg and J. Arndt. Terror Management Theory. In *The Handbook of Theories of Social Psychology*. 2012.