



University of Pittsburgh

LINGUISTIC REFLECTIONS OF SYMPTOM CONTROLLABILITY FOR WOMEN WITH RECURRENT OVARIAN CANCER

DS4LING CLASS PROJECT PRESENTATION

Yan Wang

University of Pittsburgh

Dec 6, 2021





PRESENTATION OUTLINE

Background

Methods

Results

Conclusion

Implications

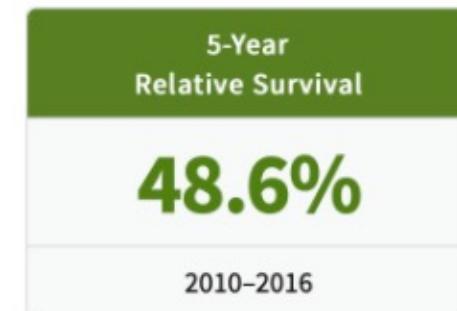
Limitations



Background – Recurrent Ovarian Cancer

At a Glance

Estimated New Cases in 2020	21,750
% of All New Cancer Cases	1.2%
Estimated Deaths in 2020	13,940
% of All Cancer Deaths	2.3%



SEER Cancer Statistics Factsheets: Ovarian Cancer.
National Cancer Institute. Bethesda, MD

Around 85% will have recurrence



Background –Problem

On average, women receive 4 different lines of chemotherapy for advanced disease; complex and aggressive treatment regimens can lead to multiple and severe symptoms (e.g., pain, neuropathy, chemo brain)

The process of trying to manage multiple symptoms can be an overwhelming and time-consuming task for both patients and health care providers (HCPs).



Background – WRITE Symptoms

Working to improve life for women with recurrent ovarian cancer

PARTICIPANT LOGIN

Welcome to the
WRITE Symptoms (GOG-259)

Research Study Homepage

WRITE Symptoms (*Written Representational Intervention To Ease Symptoms*) is a web-based symptom management program for women with recurrent ovarian cancer. The program is designed to help women find new ways to get better control over their symptoms. WRITE Symptoms is offered in two different formats: (1) nurse-delivered (through interactions with a study nurse) and (2) self-directed (through an interactive computer module).

The goal of the WRITE Symptoms Study is to find out whether these educational programs can improve symptoms and quality of life for women with recurrent ovarian cancer.

This study is an approved Gynecologic Oncology Group trial and is supported by the National Cancer Institute and the National Institute of Nursing Research (RO1NR010735).

GOG -259

Nurse-delivered WRITE Symptoms vs. Self-directed WRITE Symptoms vs. Care as Usual for optimal symptom management for women with recurrent ovarian, fallopian tube, or primary peritoneal cancer

To learn more about the Gynecologic Oncology Group visit

WRITE Symptoms was a 3-arm RCT (NR001735) of two web-based psychoeducation interventions to help recurrent ovarian cancer patients find new ways to get better control over their symptoms.

Intervention period was 8 wks.



Background – SD WRITE

Self-directed arm: an interactive web-based intervention guided the patient through the intervention elements on each of the 3 symptoms patient chose. Reponses were recorded verbatim.

Dependent variable: 3 items for each symptom addressing the extent to which the person believes that they can control the symptom on a 5- point Likert-type scale of 0 (strongly disagree) to 4 (strongly agree).



Findings of SD WRITE

SD WRITE groups showed significant symptom controllability increase from BL to 8-weeks (Mean change = 0.180, SE = 0.048, $p < .001$)

We observed variable symptom controllability change among participants, suggesting only portion of participants obtain the benefits of the intervention.



Delving into controllability

To understand the psychological processes underlying perceived symptom controllability change and potential help clinicians identify the linguistic signals, I am inspired to explore the potential linguistic features that are associated with patient perceived control.



Methods -Data preparation and analysis

- Merge & reshape dataset

SQR: ID, time of controllability assessment, symptom sequence [S1, S2, S3] and baseline and 8 week post intervention controllability scores

OEQ : ID, symptom and answers for the questions: 1. What does the symptom make you feel like; what's the cause?; what are the patterns?, 2. "How does the symptom affect you?" 3."Have you tried anything? Is it helpful?"

Metadata: ID and sociodemographics, including age, marriage status, employment, education, race, ethnicity.



Df1

	S1	S2	S3	BLS1	BLS2	BLS3	8BLS1change	8BLS2change	8BLS3change
16	Abdominal Bloating	Hot Flashes	Depression	2.2	1.4	2.4	0.8	0.2	-0.6
19	Skin Rash	Hair Loss	Sleep Disturbances	2.0	0.4	1.6	0.2	1.0	0.4
11	Peripheral Neuropathy	Lymphedema	Anxiety	2.6	2.6	2.6	-0.8	-0.6	0.0
12	Abdominal Bloating	Fatigue	Pain	2.8	1.8	3.0	1.2	1.4	0.6
12	Fatigue	Peripheral Neuropathy	Weight Gain	3.0	3.0	3.0	0.0	0.0	0.0
.1	Hot Flashes	Dizziness	Vomiting	2.4	2.8	3.0	0.6	0.2	0.8
.1	Dizziness	Drowsiness	Peripheral Neuropathy	1.2	1.8	2.4	1.2	0.4	0.0
.4	Constipation	Memory Problems	Fatigue	3.6	3.0	3.2	-0.6	0.6	0.8
13	Peripheral Neuropathy	Associated	Sleep Disturbances	2.8	2.8	2.6	1.0	0.4	0.6
							0.4	0.2	-0.2
							0.6	0.4	0.4

{r Reshape Df1}

```
Df1_1<-Df1 %>% melt(id.vars = c("ID", "GOGID"), measure.vars =c("S1", "S2", "S3"),
variable.name = "SymptomNo", value.name = "Symptom")
Df1_2<-Df1 %>% melt(id.vars = c("ID", "GOGID"), measure.vars =c( "8BLS1change",
"8BLS2change", "8BLS3change"), variable.name = "toy", value.name = "8BLChange")
Df1_3<-Df1 %>% melt(id.vars = c("ID", "GOGID"), measure.vars =c("BLS1", "BLS2",
"BLS3"), variable.name = "toy", value.name = "BSContr")
colnames(Df1_1)
colnames(Df1_2)
colnames(Df1_3)
levels(Df1_2$toy)[1]<- "S1"
levels(Df1_2$toy)[2]<- "S2"
levels(Df1_2$toy)[3]<- "S3"
names(Df1_2)[3]<- "SymptomNo"
levels(Df1_3$toy)[1]<- "S1"
levels(Df1_3$toy)[2]<- "S2"
levels(Df1_3$toy)[3]<- "S3"
names(Df1_3)[3]<- "SymptomNo"
Df1_4<- inner_join(Df1_1, Df1_2, by= c("ID", "GOGID", "SymptomNo")) %>%
inner_join(., Df1_3, by= c("ID", "GOGID", "SymptomNo")) %>% na.omit()
```

```

## Df1\_2

|  | toy         | 8BLChange |
|--|-------------|-----------|
|  |             | <dbl>     |
|  | 8BLS1change | 0.8       |
|  | 8BLS1change | 0.2       |
|  | 8BLS1change | -0.8      |
|  | 8BLS1change | 1.2       |



## Merge with OEQ (text data) by Participant ID and Symptom

| ID        | SymptomNo | Symptom               | 8BLChange | BSContr | Feelin   |
|-----------|-----------|-----------------------|-----------|---------|----------|
| ·0259-006 | S1        | Abdominal Bloating    | 0.8       | 2.2     | I`m al   |
| ·0259-039 | S1        | Skin Rash             | 0.2       | 2.0     | Red e    |
| ·0259-001 | S1        | Peripheral Neuropathy | -0.8      | 2.6     | My sy    |
| ·0259-002 | S1        | Abdominal Bloating    | 1.2       | 2.8     | The bl   |
| ·0259-002 | S1        | Fatigue               | 0.0       | 3.0     | About    |
| ·0259-011 | S1        | Hot Flashes           | 0.6       | 2.4     | I have   |
| ·0259-011 | S1        | Dizziness             | 1.2       | 1.2     | usuall   |
| ·0259-014 | S1        | Constipation          | -0.6      | 3.6     | Straini  |
| ·0259-002 | S1        | Peripheral Neuropathy | 1.0       | 0.8     | I live v |
| ·0259-011 | S1        | Fatigue               | 0.4       | 1.8     | I`m fi   |
| ·0259-009 | S1        | Constipation          | 0.6       | 2.4     | This s   |



# Data analysis— feature extraction

```
1 symptom
2 effort
3 feel
4 impact
5 positive adj
6 negative adj
7 controlled
8 uncontrolled
%

abdominal bloating 1
bloating 1
pain 1
fatigue 1
lack of energy 1
anxiety 1
panic 1
constipation 1
depression 1
diarrhea 1
dizzy 1
dizziness 1
drowsy 1
----- 1
```

The screenshot shows two windows side-by-side. The top window is the LIWC software, titled 'Welcome to LIWC'. It features four main sections: 'Analyze Text' (described as 'See how text measures up along LIWC's dimensions'), 'Categorize Words' (described as 'Break down each word and find the categories it applies to'), 'Color-Code Text' (described as 'Highlight words found in LIWC's dictionary'), and 'Category Options' (described as 'Choose which dimensions to use during analysis'). Below these sections are 'Help' and 'Settings' tabs. The settings tab shows the following configuration: 'Dictionary' set to 'Internal Dictionary 2015', 'Segmentation' set to 'No segmentation', and 'All categories on' set to 'NO'. There is also a checked checkbox for 'Show this window when LIWC opens'. The bottom window is the 'LightSide' researcher's workbench. It has a toolbar with tabs for 'Extract Features', 'Restructure Data', 'Build Models', 'Explore Results', 'Compare Models', and 'Predict Labels'. The main area contains several input fields: 'CSV Files' (with a browse button), 'Feature Extractor Plugins' (checkboxes for 'Basic Features', 'Character N-Grams', 'Column Features', 'English Parse Features', 'Regular Expressions', and 'Stretchy Patterns', with 'Basic Features' checked), 'Configure Basic Features' (checkboxes for 'Unigrams', 'Bigrams', 'Trigrams', 'POS Bigrams', 'POS Trigrams', 'Word/POS Pairs', 'Line Length', 'Count Occurrences', 'Normalize N-Gram Counts', 'Include Punctuation' (checked), 'Stem N-Grams', and 'Skip Stopwords in N-Grams', with 'Include Punctuation' checked), 'Text Fields' (input field with a 'Differentiate Text Fields' checkbox), and buttons for 'Extract', 'Name: features', and 'Rare Threshold: 5'.



## Data analysis

Linear mixed effects model

Random effects: participant ID

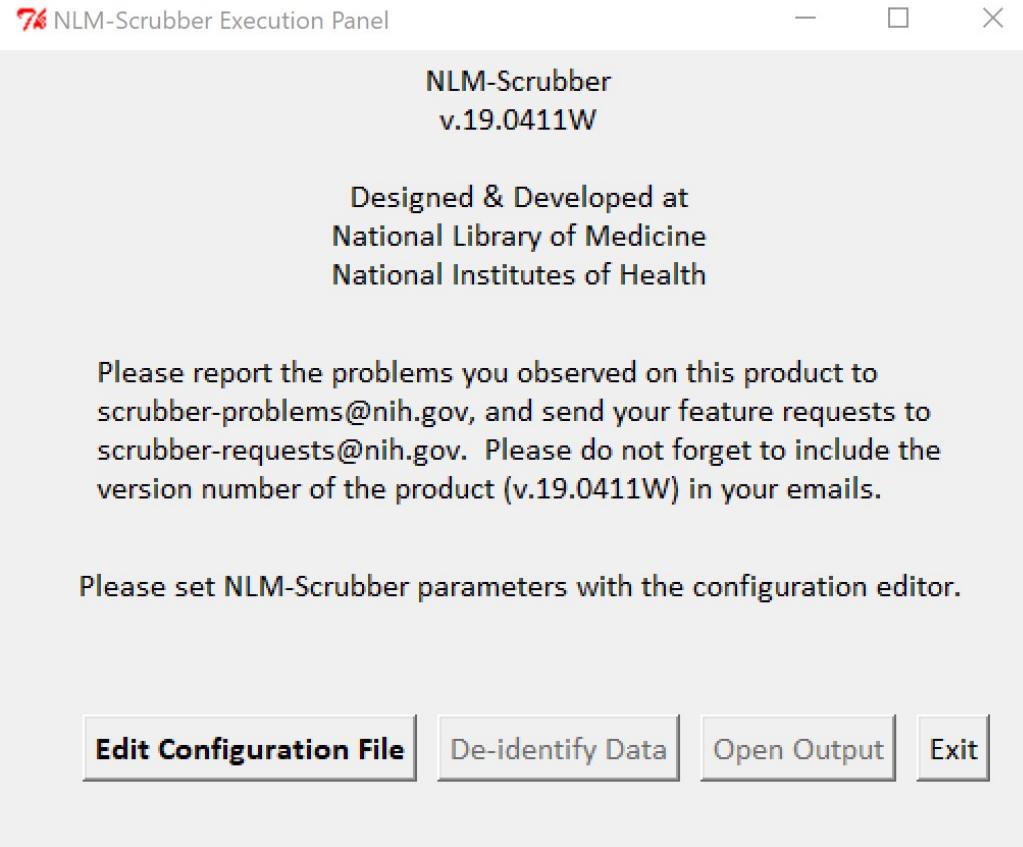
Covariates: participant social demographic factors, e.g.,  
marriage status, age, education, employment

Fixed effects: baseline symptom controllability score and  
individual linguistic features or new predictors created by PCA  
on all the individual linguistic features



# Data sharing plan

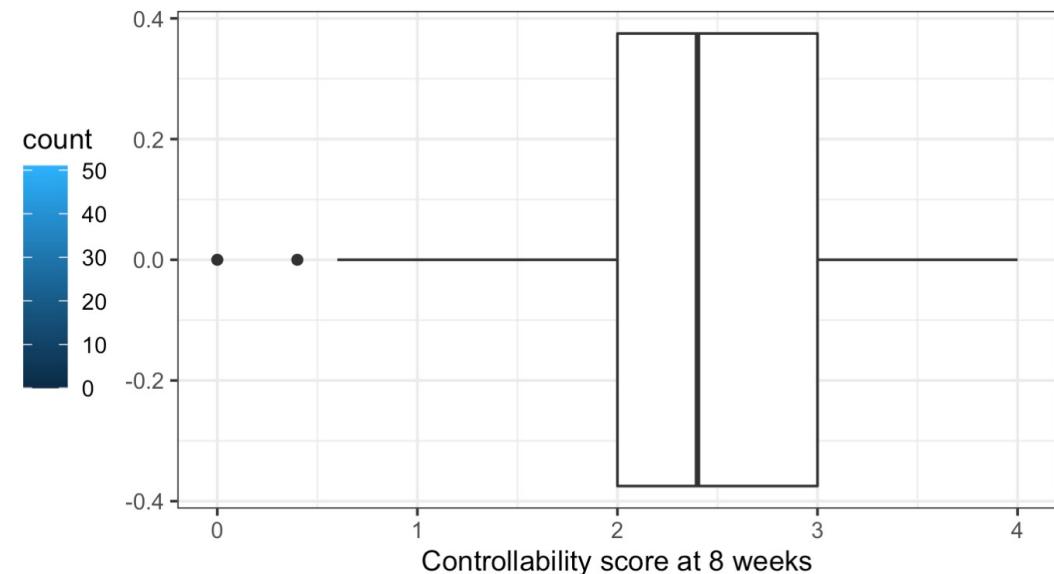
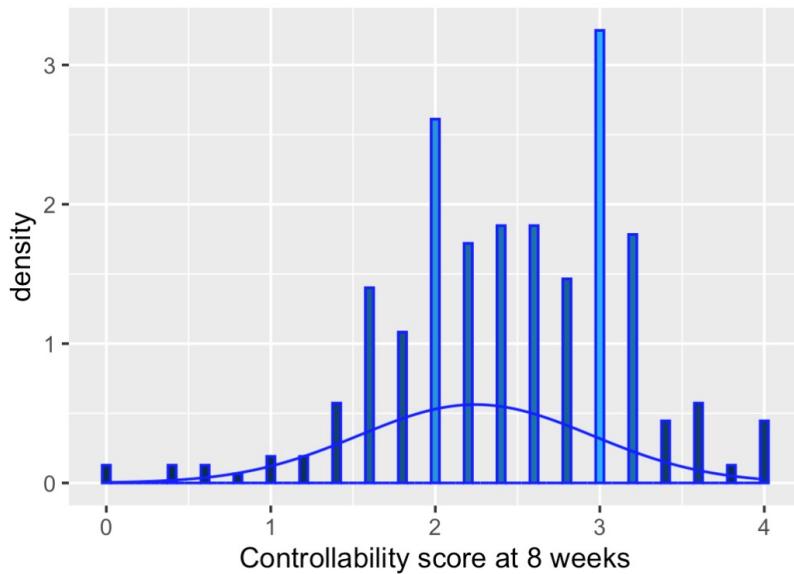
NLM-Scrubber





# Results

## Perceived controllability @ 8 weeks



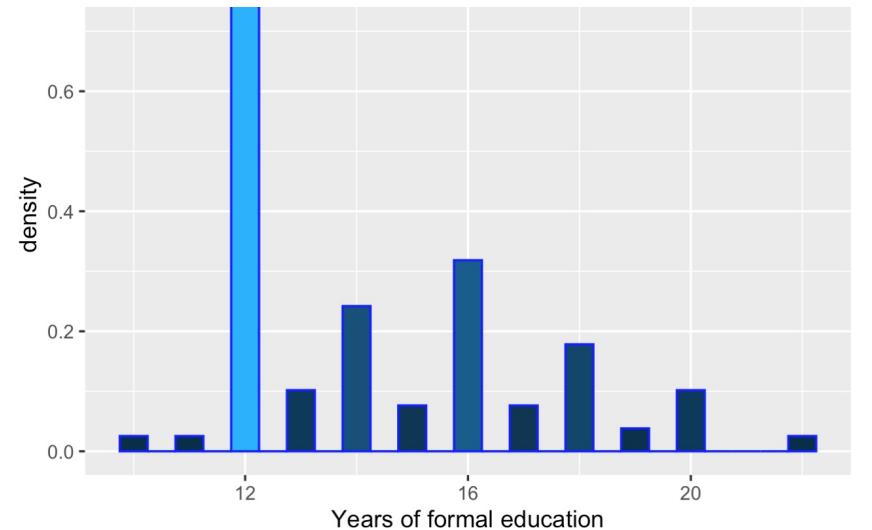
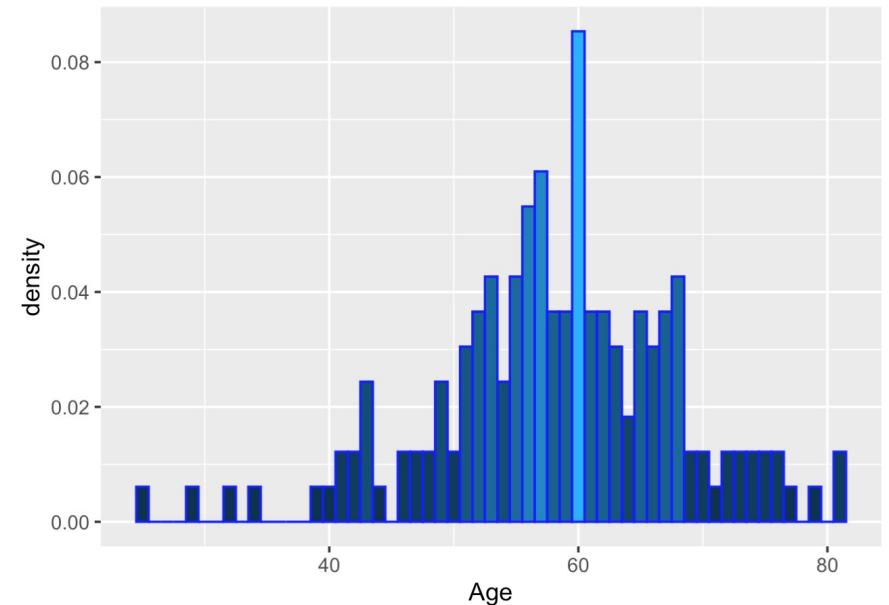


## Participant sample

Predominately non-Hispanic(96.18%), White (93%), older(75.6%), and married or living with partner (75.16%)

Over half is employed (59.24%)

The mean of age is 58.18 (SD = 9.72). The average of formal years of education is 14.4 (SD=2.72)





# 1. ID as random effects and individual linguistic features as predictors with opinionated forward selection

Random effects:

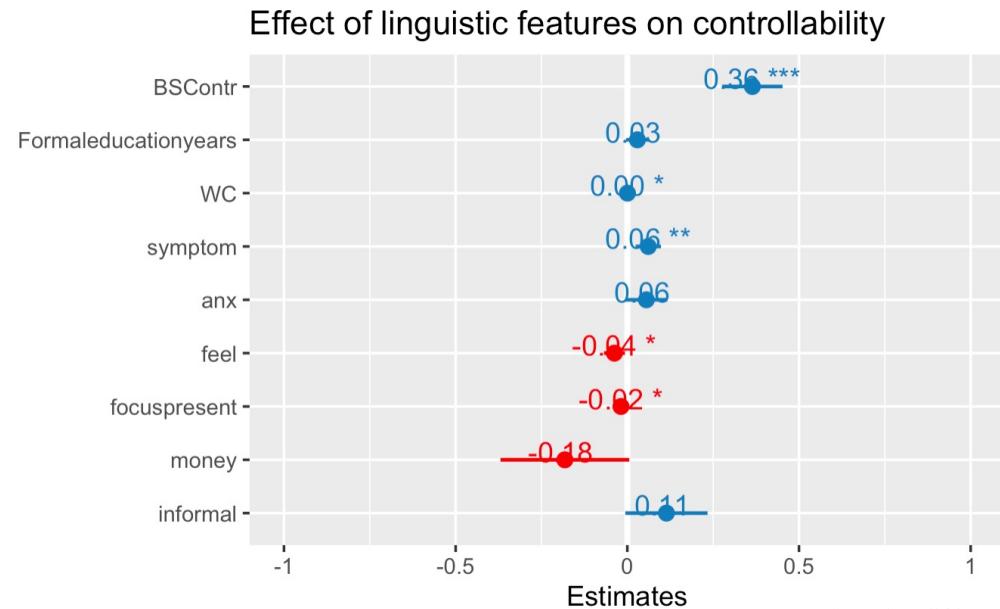
| Groups | Name        | Variance | Std.Dev. |
|--------|-------------|----------|----------|
| ID     | (Intercept) | 0.1723   | 0.4151   |
|        | Residual    | 0.1774   | 0.4212   |

Number of obs: 300, groups: ID, 114

Fixed effects:

|                      | Estimate   | Std. Error | df        | t value | Pr(> t )     |  |
|----------------------|------------|------------|-----------|---------|--------------|--|
| (Intercept)          | 1.313e+00  | 3.149e-01  | 1.926e+02 | 4.169   | 4.62e-05 *** |  |
| BSContr              | 3.642e-01  | 4.470e-02  | 2.789e+02 | 8.147   | 1.26e-14 *** |  |
| Formaleducationyears | 2.934e-02  | 1.678e-02  | 1.088e+02 | 1.748   | 0.08322 .    |  |
| WC                   | 4.199e-04  | 1.724e-04  | 2.892e+02 | 2.436   | 0.01545 *    |  |
| symptom              | 6.103e-02  | 1.879e-02  | 2.384e+02 | 3.248   | 0.00133 **   |  |
| anx                  | 5.555e-02  | 3.106e-02  | 2.201e+02 | 1.788   | 0.07511 .    |  |
| feel                 | -3.761e-02 | 1.561e-02  | 2.398e+02 | -2.409  | 0.01675 *    |  |
| focuspresent         | -1.849e-02 | 9.377e-03  | 2.627e+02 | -1.972  | 0.04968 *    |  |
| money                | -1.819e-01 | 9.525e-02  | 2.524e+02 | -1.909  | 0.05737 .    |  |
| informal             | 1.139e-01  | 6.085e-02  | 2.288e+02 | 1.872   | 0.06244 .    |  |

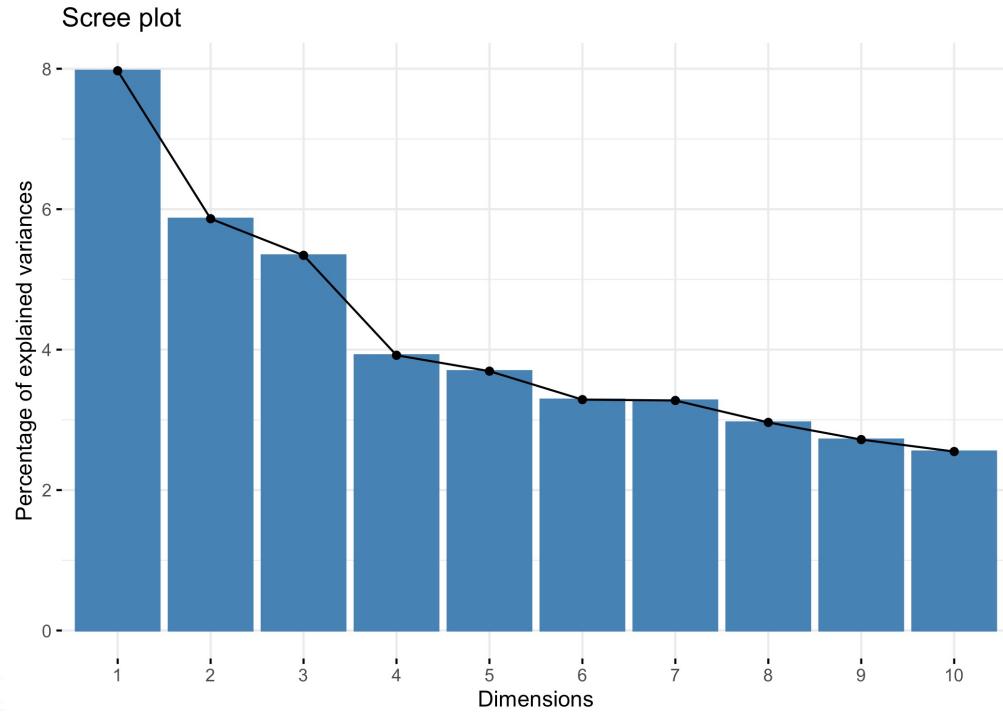
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1



| Term                 | GVIF     | Df | GVIF^(1/(2*Df)) | GVIF^(1/Df) |
|----------------------|----------|----|-----------------|-------------|
| BSContr              | 1.034703 | 1  | 1.017204        | 1.034703    |
| Formaleducationyears | 1.027945 | 1  | 1.013876        | 1.027945    |
| WC                   | 1.089354 | 1  | 1.043721        | 1.089354    |
| symptom              | 1.115491 | 1  | 1.056168        | 1.115491    |
| anx                  | 1.048245 | 1  | 1.023838        | 1.048245    |
| feel                 | 1.055237 | 1  | 1.027248        | 1.055237    |
| focuspresent         | 1.051768 | 1  | 1.025557        | 1.051768    |
| money                | 1.061247 | 1  | 1.030169        | 1.061247    |
| informal             | 1.028593 | 1  | 1.014196        | 1.028593    |



## 2.1 PCA on all the linguistic features



33 components have eigen values  $>1$ , explaining 76.8% of the variance in total



## 2.2 ID as random effects, principle components are fixed effects with iterative backward elimination

Random effects:

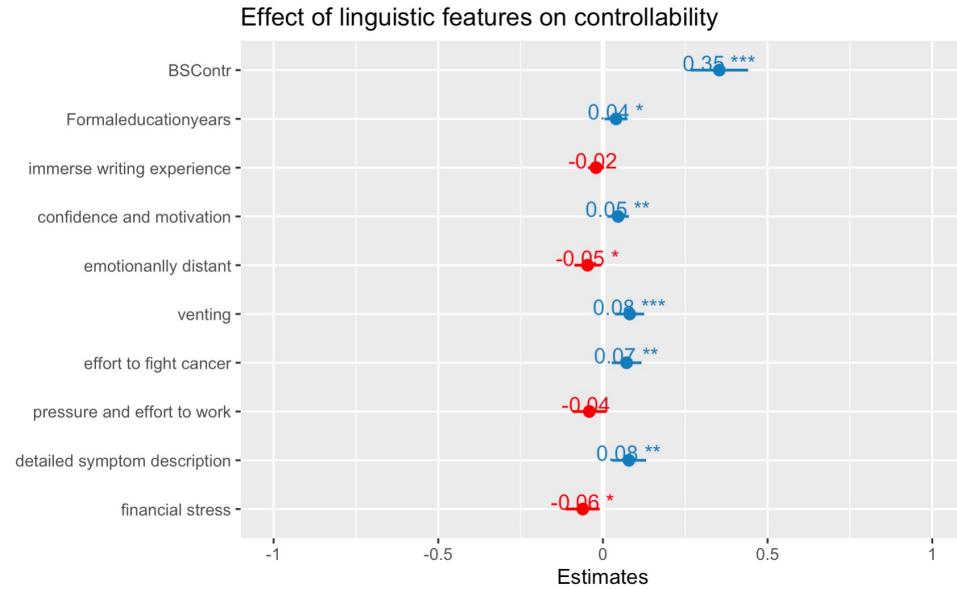
| Groups   | Name        | Variance | Std.Dev. |
|----------|-------------|----------|----------|
| ID       | (Intercept) | 0.2148   | 0.4634   |
| Residual |             | 0.1546   | 0.3931   |

Number of obs: 300, groups: ID, 114

Fixed effects:

|                              | Estimate | Std. Error | df        | t value | Pr(> t )     |
|------------------------------|----------|------------|-----------|---------|--------------|
| (Intercept)                  | 1.09827  | 0.27862    | 125.43539 | 3.942   | 0.000134 *** |
| BSContr                      | 0.35347  | 0.04466    | 274.05592 | 7.914   | 6.2e-14 ***  |
| Formaleducationyears         | 0.03984  | 0.01784    | 109.01921 | 2.234   | 0.027540 *   |
| immerse.writing.experience   | -0.02095 | 0.01255    | 240.95340 | -1.670  | 0.096285 .   |
| confidence.and.motivation    | 0.04615  | 0.01681    | 258.02543 | 2.745   | 0.006481 **  |
| emotionally.distant          | -0.04653 | 0.02053    | 258.79106 | -2.266  | 0.024283 *   |
| venting                      | 0.08141  | 0.02232    | 218.71358 | 3.647   | 0.000332 *** |
| effort.to.fight.cancer       | 0.07200  | 0.02311    | 217.49177 | 3.115   | 0.002086 **  |
| pressure.and.effort.to.work  | -0.04140 | 0.02575    | 247.57830 | -1.608  | 0.109115     |
| detailed.symptom.description | 0.07900  | 0.02662    | 236.15700 | 2.968   | 0.003302 **  |
| financial.stress             | -0.06117 | 0.02626    | 217.72126 | -2.329  | 0.020761 *   |

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

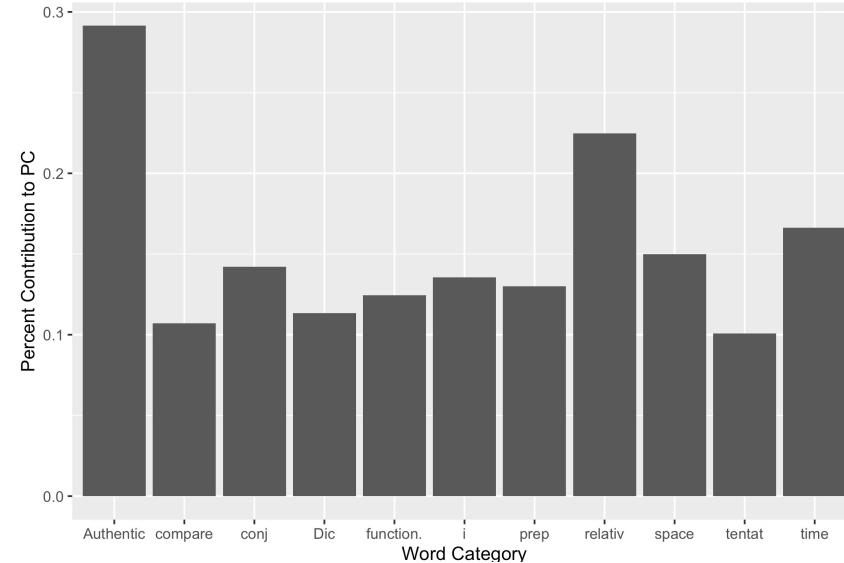


| Term                         | GVIF     | Df | GVIF^(1/(2*Df)) | GVIF^(1/Df) |
|------------------------------|----------|----|-----------------|-------------|
| BSContr                      | 1.105753 | 1  | 1.051548        | 1.105753    |
| Formaleducationyears         | 1.020740 | 1  | 1.010317        | 1.020740    |
| immerse.writing.experience   | 1.025309 | 1  | 1.012576        | 1.025309    |
| confidence.and.motivation    | 1.021201 | 1  | 1.010545        | 1.021201    |
| emotionally.distant          | 1.062346 | 1  | 1.030702        | 1.062346    |
| venting                      | 1.014000 | 1  | 1.006976        | 1.014000    |
| effort.to.fight.cancer       | 1.022894 | 1  | 1.011382        | 1.022894    |
| pressure.and.effort.to.work  | 1.036182 | 1  | 1.017930        | 1.036182    |
| detailed.symptom.description | 1.014454 | 1  | 1.007201        | 1.014454    |
| financial.stress             | 1.006390 | 1  | 1.003190        | 1.006390    |

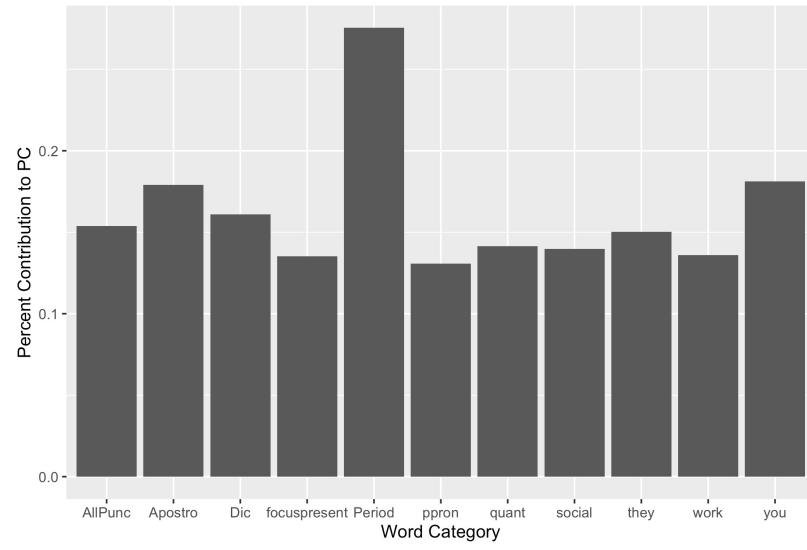


## 2.2 Contribution to the principle components

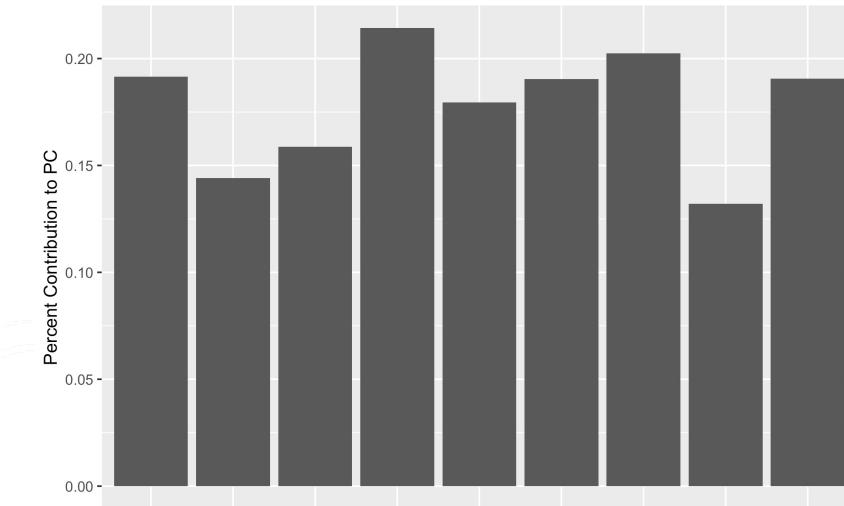
PC3:Immersive Writing



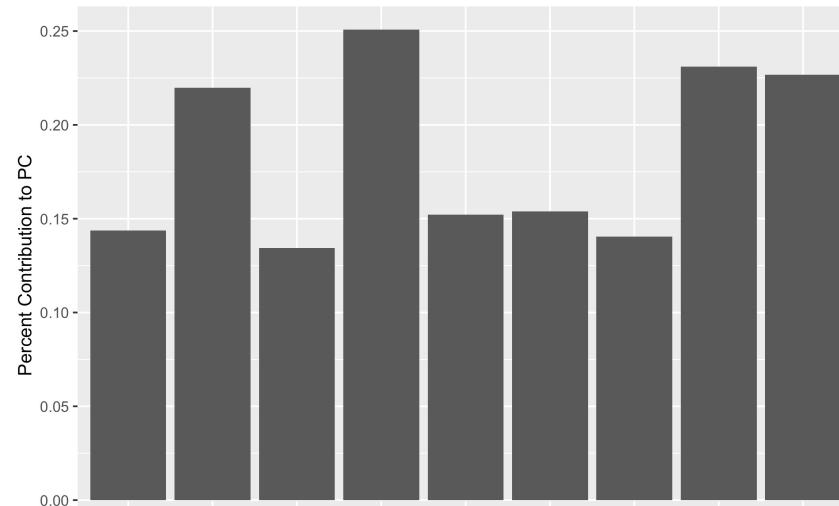
PC13:Emotionally distant



PC7:Confidence and motivation



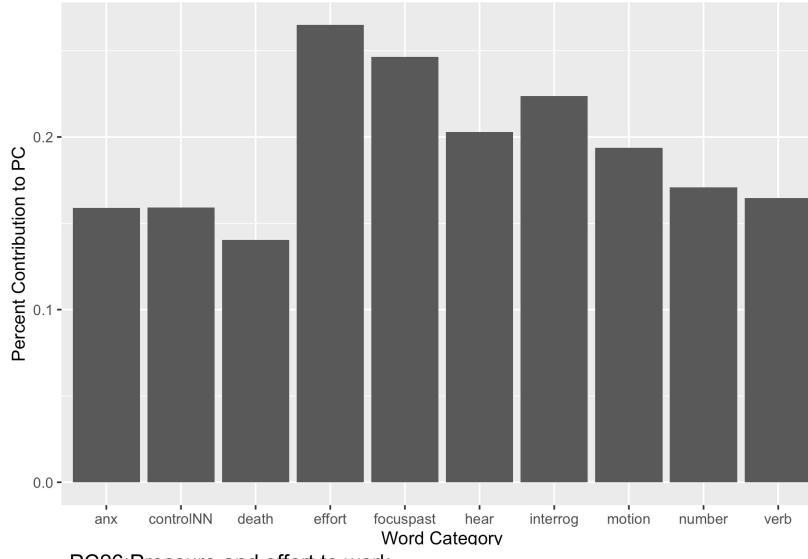
PC22:Venting(confusion and frustration)



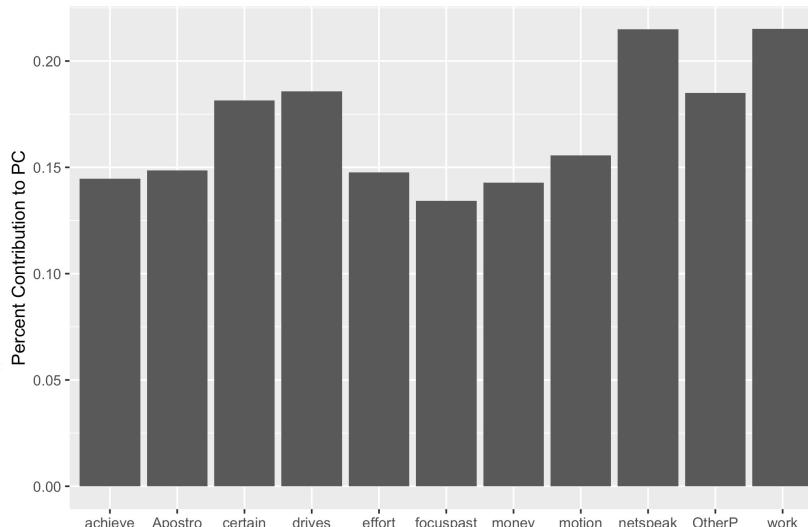


## 2.2 Contribution to the principle components

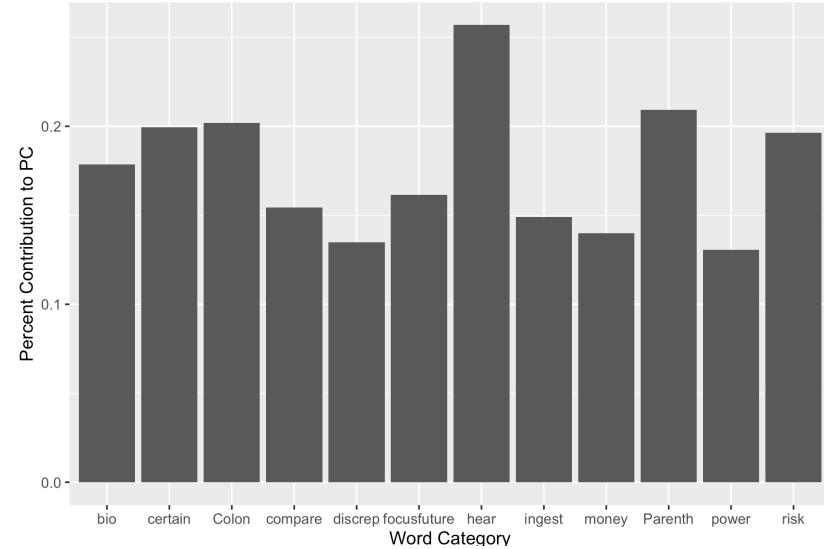
PC24:Effort to fight against cancer



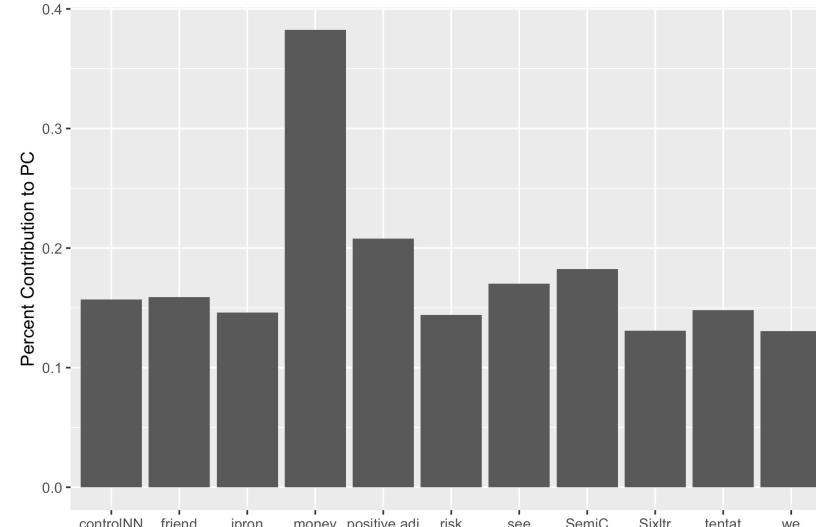
PC26:Pressure and effort to work



PC30:detailed.symptom.description



PC33:financial.stress



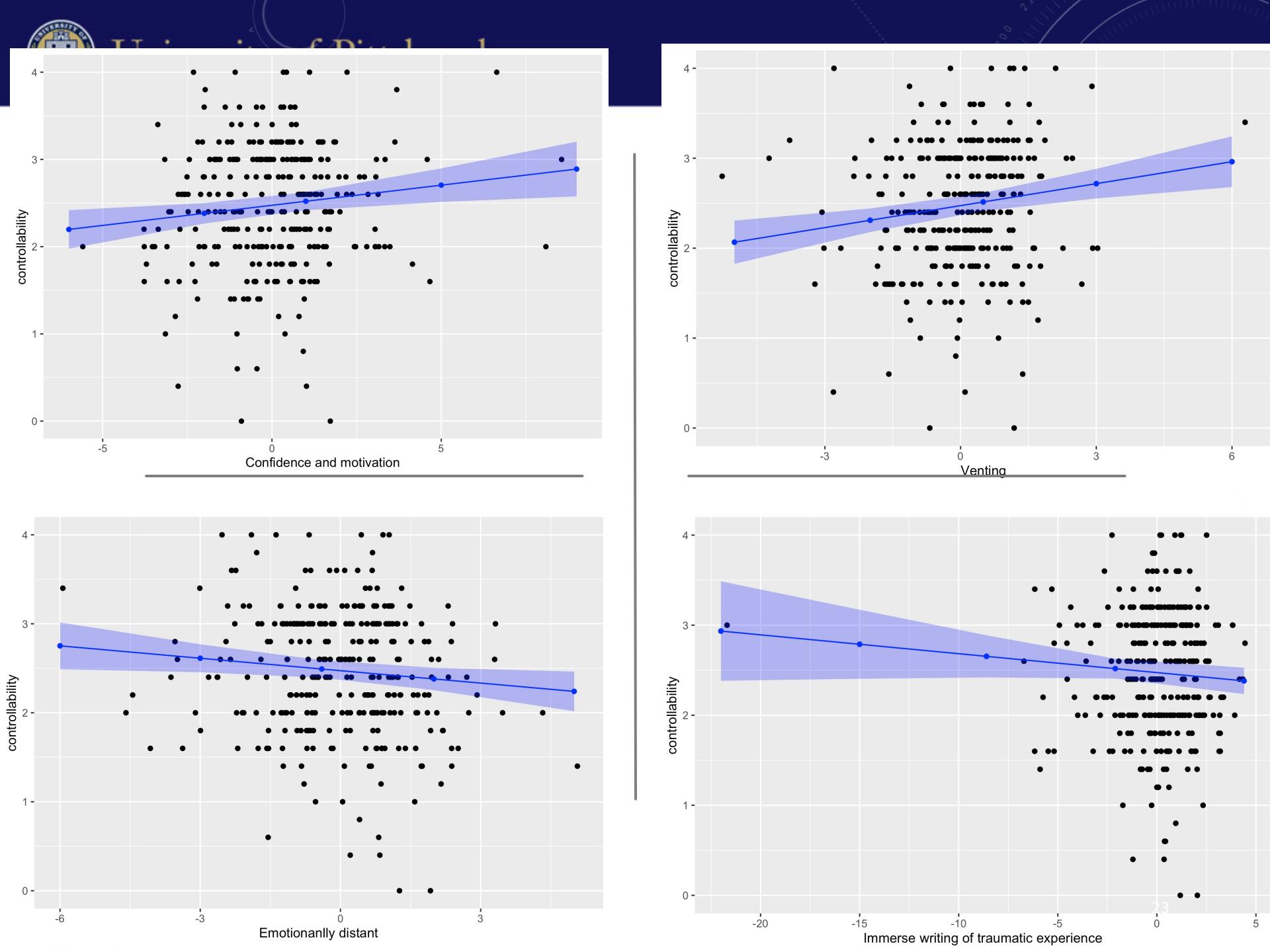


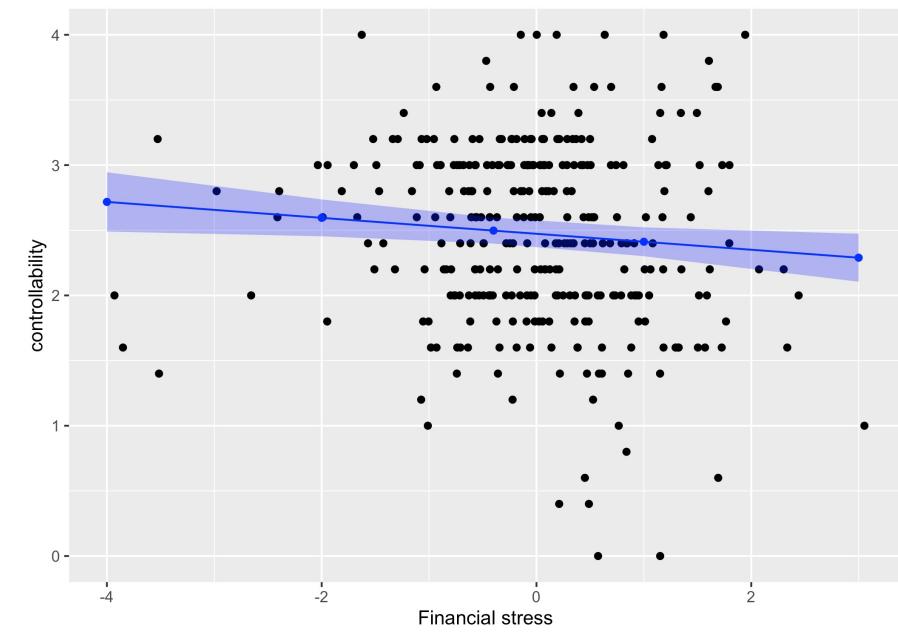
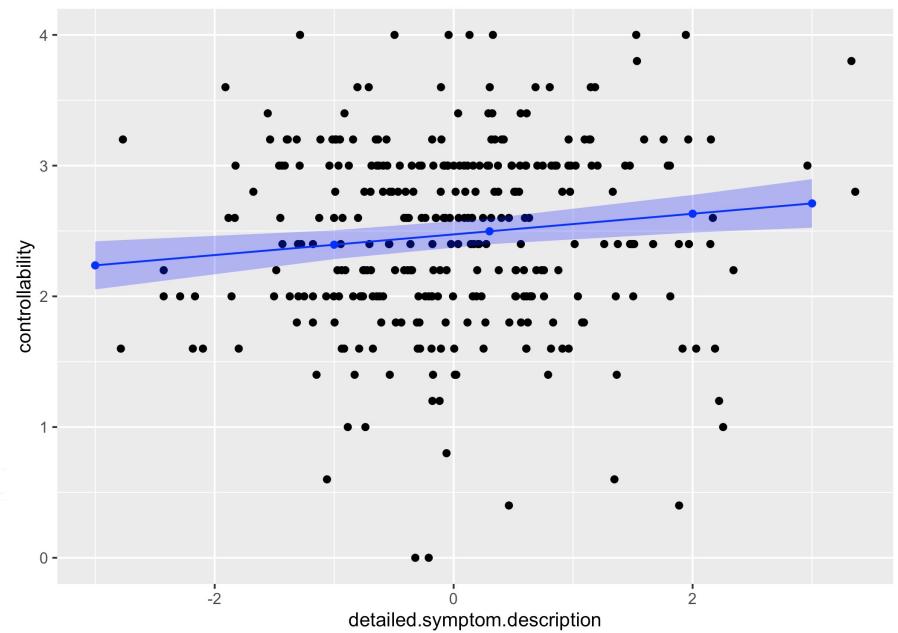
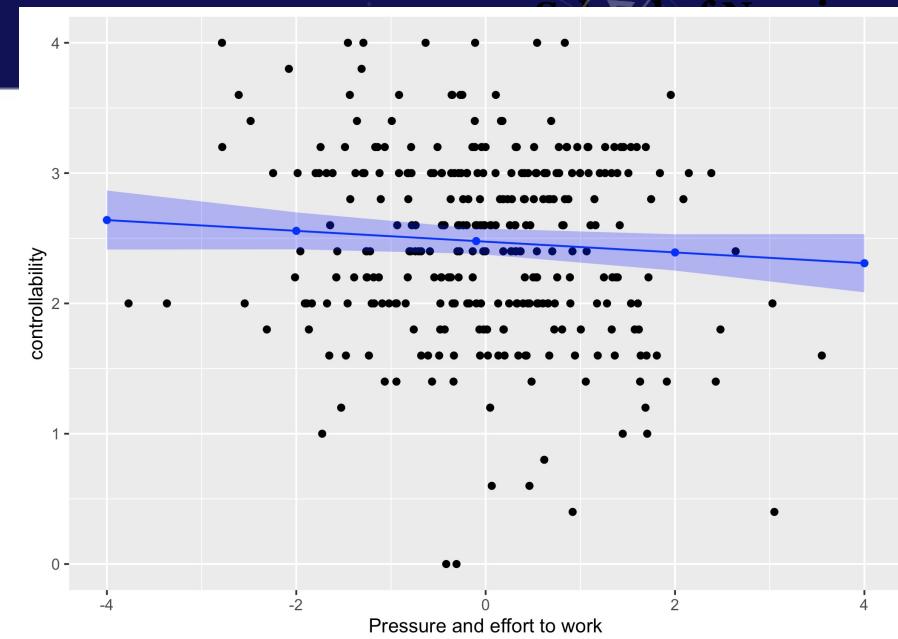
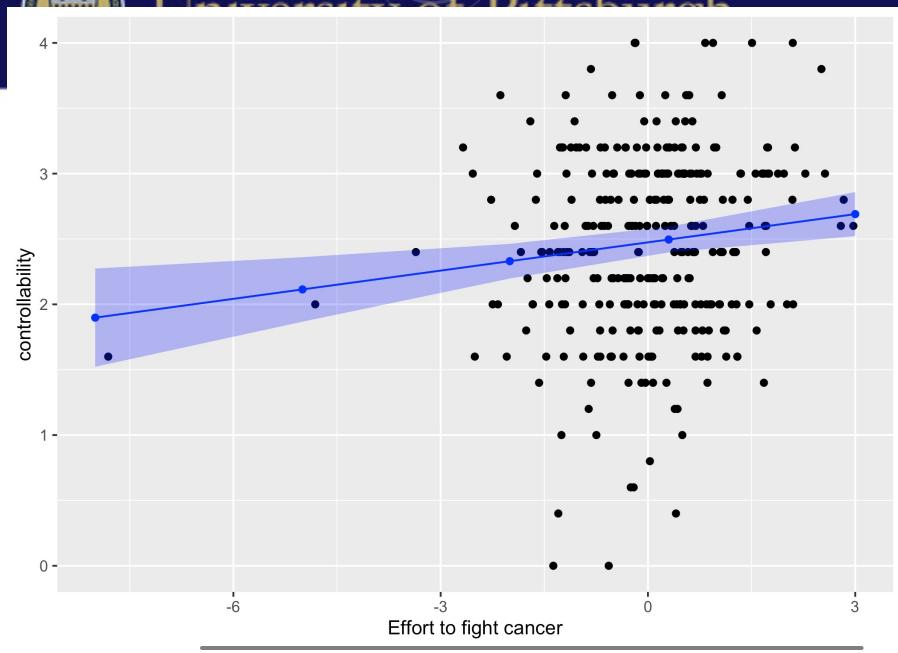
### 3. Compare two models

|       | <b>df</b><br><code>&lt;dbl&gt;</code> | <b>AIC</b><br><code>&lt;dbl&gt;</code> |
|-------|---------------------------------------|----------------------------------------|
| toy13 | 13                                    | 540.2030                               |
| toy11 | 12                                    | 551.1619                               |

|       | <b>df</b><br><code>&lt;dbl&gt;</code> | <b>BIC</b><br><code>&lt;dbl&gt;</code> |
|-------|---------------------------------------|----------------------------------------|
| toy13 | 13                                    | 588.3522                               |
| toy11 | 12                                    | 595.6073                               |

Seems like model toy 13 is better







## Conclusions

Based on the controllability linguistic markers, it is important for the patients to reflect their symptoms in a personal and meaningful way although there might be risk that immerse writing can evoke negative emotions. When controlled for participants and their baseline controllability scores and education, patients who are confident and motivated, tend to release negative emotions, and make great efforts to fight cancer in general are more likely to have better control of their symptoms. Those who are experiencing stress from employment and finance are more likely to do poorly in symptom control.



# Implications

In contrast to f2f intervention where patients engage with materials in a structured and monitored way, and clinicians directly observe patients behavior and provide feedback, in digital intervention, it is important to target the limited clinician's attention to patients who need it most.

The linguistic features identified in the message board can help clinicians monitor patient sense of control. By identifying patients who are likely to end up not improve their ability to control symptoms before it is too late, the clinicians can provide emotional support and tailor the content in time to help those patients control their symptom better.