

Cancer Survivorship Analysis

A Longitudinal Quality of Life Analysis of Cancer Survivors

Riley Waters

University of Washington
Seattle, WA

Adrian Tullock

University of Washington
Seattle, WA

Hallie Ertman

University of Washington
Seattle, WA

Naman Sharma

University of Washington
Seattle, WA
in.namansharma@gmail.com

Abstract

Survival rates of cancer have improved drastically over the years, but many of its effects continue to influence the lives of survivors well past their diagnosis. Fred Hutchinson Cancer Center has monitored these effects through an extensive surveying and diagnosis collection system called SIMS. In this project, we reviewed and analyzed the SIMS dataset, which includes self-reported quality of life metrics and clinical information on hundreds of cancer survivors. We explored the dataset using traditional inference techniques to find patterns in responses between baseline and follow-up surveys. We also employed regression analysis to further explore individual correlations between features in the survey data. Random forest models were used to see if predictions of quality of life indicators could be accurately made given the rest of the survey data.

In the longitudinal analysis, we find that patients are largely walking away from their treatment educated about their plans and personal health, and are not experiencing notable detriments to their perceived quality of life between the baseline and follow-up surveys. Many well-established correlations were confirmed in regression analysis, but we also found that some domains of predictor variables consistently failed to correlate with quality of life indicators. Predictions using random forest models were proven accurate, but the models also relied heavily on small sets of intuitively correlated variables. The significant amount of work that has gone into creating the data pipeline, data mappings, and reproducible analyses will be useful tools in generating more key insights from the SIMS data. That said, there are findings presented that yield valuable information about the survivorship program.

Keywords—cancer; analysis; survivorship; machine learning;

I. INTRODUCTION

Fred Hutchinson Cancer Research Center has created a centralized cancer survivorship data repository called SIMS (Survivorship Information Management System). The system compiles survey data from survivors collected prior to their first clinic visit and during annual follow-ups. The dataset has fields pertaining to demographics, lifestyle, medications, pain, sleep and more. There is also clinical diagnosis data containing medical information. In total, the data has over 3500 columns and 950 rows of unique patients. This project aims to analyze the SIMS dataset for trends that can bolster our understanding of the long-term effects of cancer, characteristics of patients, and actionable causes that may aid patients in the clinical process.

II. PROBLEM STATEMENT

Clinicians actively use SIMS for patient evaluation needs, but no major statistical analyses have been performed on it. The purpose of this project is to gain analytical insights that may improve the clinical process for cancer patients and better understand how cancer affects their lives. Specifically, we addressed the following questions:

Capstone Project for Masters in Data Science at University of Washington
March 2020

- *Can quality of life indicators be accurately predicted using the survey data?*
- *What associations can be found between patient characteristics and health behaviors?*
- *What common trends exist when comparing the baseline and follow-up surveys?*

III. BACKGROUND

Cancer survivorship encompasses numerous other aspects that factor into the long-term experience of survivors. The term aftereffects defines long-term or late effects of cancer and treatment. While long-term effects surface during treatment and can continue post-treatment, late effects can surface months to years after treatment ends. This includes effects that are assuaged over time and chronic in nature [1]. Knowing this, depending on the longevity of the data we have, there may be little we can contribute to knowledge on late effects.

The range of effects are also related to the type of cancer and treatment a patient experiences. Table 1 provides a list of examples by treatment type. For instance, radiation, chemotherapy, and surgery are all separate solutions. Survivor physical, mental, and emotional health consequences can be attributed to different sources due to this. Whereas physical scars are not uncommon with surgical procedures, chemotherapy can impair cell growth and reproduction. Similarly, breast cancer and prostate cancer are also separate health issues [2].

TABLE I.

Possible After-effects Associated with...		
Chemotherapy	Radiation Therapy	Surgery
<ul style="list-style-type: none">• Changes in hair or nails• Cognitive problems• Fatigue• Memory loss• Muscle weakness• Numbness, tingling	<ul style="list-style-type: none">• Cataracts• Dental decay, tooth loss• Inability to produce tears/saliva• Infertility• Secondary cancers• Skin sensitivity to sun exposure	<ul style="list-style-type: none">• Changes in fertility• Changes in sexual function• Cognitive problems• Internal/External scarring• Problems with movement• Swelling of arms or legs

Table 1. A non-exhaustive list of after effects associated with treatment types

Along with implications on day-to-day life and biological functions, body image, stress, general uncertainty, and other factors can challenge the emotional health of cancer survivors. Some bodily changes, such as hair loss or weight change, can be temporary and/or addressed by lifestyle adjustments. Others, such as scars, amputations, or infertility, are permanent. Negative effects surrounding body image issues are not solely a matter of the visibility of the change. Self-esteem, feelings of insecurity and uncertainty, or mental perceptions of differences all can affect a survivor's quality of life [3].

Some known sources of stress include feelings of isolation, preoccupying thoughts of cancer, and trouble finding a new sense of normalcy. Trouble adjusting to life post-treatment has been observed to promote sadness, and even depression, among survivors. Although, some level of sadness is expected overall when dealing with cancer, in general, feelings of sadness often do not last very long, albeit they can be strong. Mild depression can set in if the issue goes on for a prolonged period of time [4]. This can be identified when individuals are able to function normally, but only with higher-than-normal levels of exertion. Beyond that, when the sad feelings become overwhelming and everyday activities become very difficult, clinical depression can set in. Some signs of clinical depression may include a loss of interest in activities that were once enjoyed, fatigue, problems with memory and concentration, feelings of hopelessness, or body aches [5].

Adapting to life as a survivor carries a varying range of challenges from one person to the next. However, there are some general suggestions for coping with some categorical obstacles. For problems with stress and mood, including sadness and depression, attending support groups, talking with family and friends, reading, meditation, prayer, and breathing exercises are common recommendations [6]. Other recommendations that have added applications in aiding body image concerns and overall health, engaging in regular physical activity and exercise, as well as adapting a healthy diet [7].

IV. DATA SUMMARY

The data is an aggregate of multiple sources including past clinical information, patient survey responses,

Capstone Project for Masters in Data Science at University of Washington March 2020

and follow-up responses. The clinical information contains ~1200 rows of data and 1400+ columns. The initial survey responses number ~700 rows and ~640 columns in total. The follow-up responses have ~430 rows and ~270 columns. The data has been filtered of private identifiers, compiled together, and expanded in the case of categorical responses. The final source has over 3500 columns and at least 950 rows of unique patients.

The data sits in Fred Hutch-controlled, on-premise file management systems. The infrastructure team has built proprietary data management systems for access control over the data. These systems also include extensive scientific computing resources to perform operations over the protected data.

There are thousands of features in our datasets that may prove useful, but we have identified the following fields that are of interest to our questions:

- *Demographics*: Race, income, insurance, employment, occupation (Many of these fields were removed in de-identification but some subcategories may still be present)
- *Fatigue*: Level of fatigue, impact on daily life, how it has changed over time, impact on mood
- *General health*: Different body parts, diseases, injuries
- *Neuropathy*: discomfort and numbness
- *Pain*: General pain, impact on mood, impact on walking
- *Reproductive health*: Sexual activity and sexual functioning
- *Sleep*: time spent sleeping, difficulties
- *Health Care Visits*: Number of overnight hospital stays in the past year, number of emergency visits in the past year, various kinds of specialty care obtained since the end of cancer treatment
- *Lifestyles*: Smoking, social relationships, diet, exercise
- *Medications*: medications for depression, anxiety, panic
- *Knowledge*: Patient's confidence in his or her own knowledge about different aspects of treatment
- *Mood and worries*: losing "myself" in all the changes, dealing with insurance, appreciating life more because of having had cancer

- *General Quality of Life*: Limitations caused by health, change over time, feeling full of life
- *Muscles and joints*: Cramps and spasms, Muscle Weakness, Joint replacements, Joint and spine problems
- *Clinical Diagnostic Information*: Diagnosis age, diagnosis date, surgeries, transplants, and many medical readings

Assessment of the demographic split was done through visualizations (Figure 1-5), to ascertain a healthy split of population.

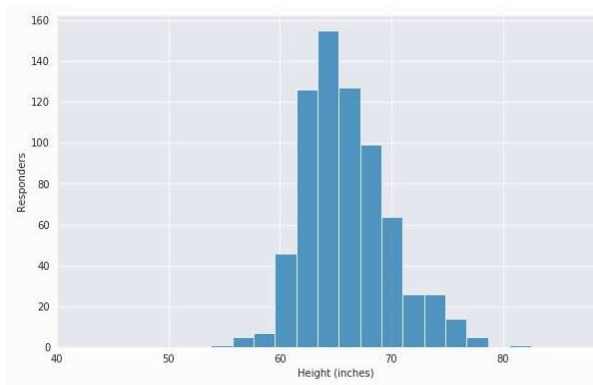


FIGURE 1. POPULATION HEIGHT DISTRIBUTION

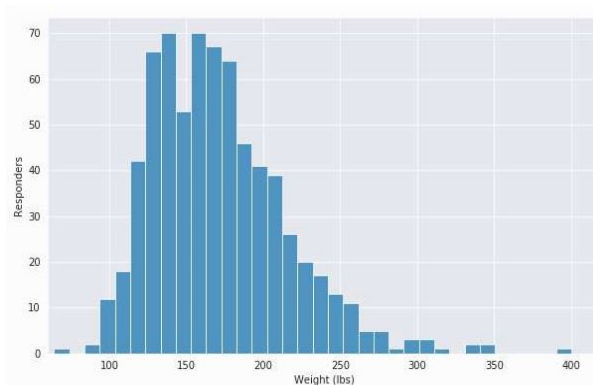


FIGURE 2. POPULATION WEIGHT DISTRIBUTION

Capstone Project for Masters in Data Science at University of Washington

March 2020

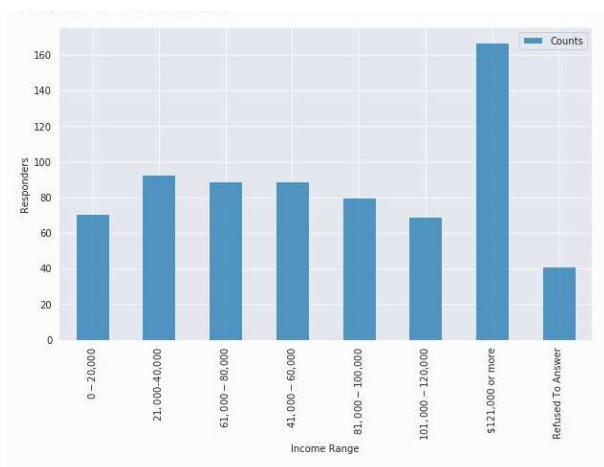


FIGURE 3. INCOME DISTRIBUTION

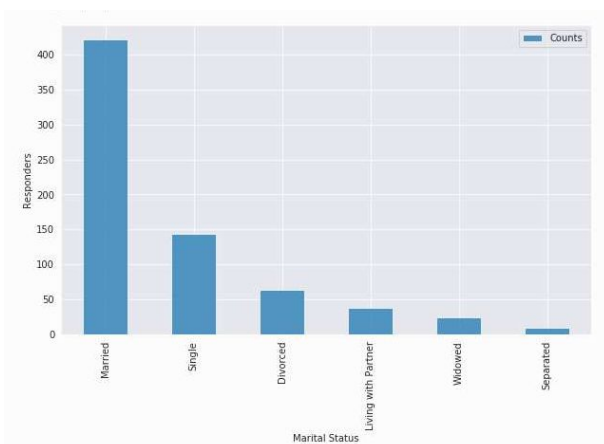


FIGURE 4. MARITAL STATUS

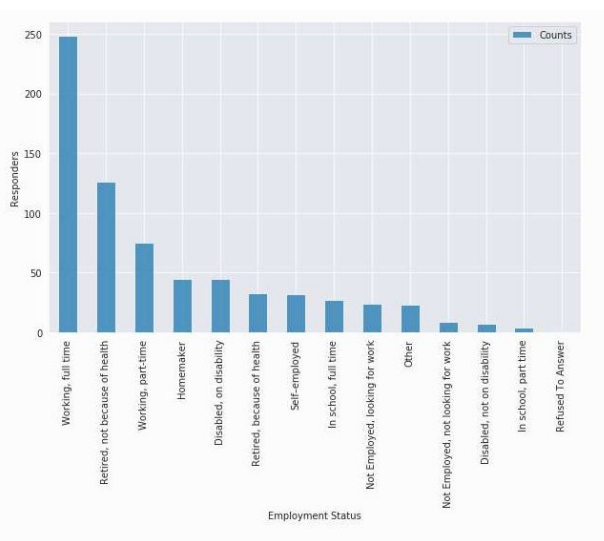


FIGURE 5. EMPLOYMENT STATUS

V. APPROACH

There were multiple approaches taken simultaneously in order to address the various limitations and opportunities present in the data. The most potential features were present in the clinical data, yet these would also be the most difficult to develop a value proposition for given the lack of specialized knowledge at hand. Within the survey data, there were opportunities to apply predictive models on responses for each individual survey type (baseline and follow-up), as well as each section within the surveys. There was also the avenue of doing longitudinal analysis for respondents of both surveys. Although, the obstacle here would be pairing the responses across surveys, which would stand as a significant deliverable all on its own. Also, the dataset for this style of analysis would afford the smallest working set, as it would be limited to those who completed both surveys with an upper limit of respondents who completed the follow-up (the smallest raw dataset).

Once the majority of the data exploration and cleaning had been completed, the sponsors expressed the value of a longitudinal analysis to be priority, along with having any analysis and findings being presented in a reproducible fashion. However, given that performing a longitudinal analysis would only require the survey data, and would not require the entire team to complete, we mobilized on multiple fronts, as to offer a wider breadth of future research paths, as opposed to an approach geared towards depth of any particular analysis.

The most challenging aspect of the longitudinal analysis would be building out a pipeline to match responses across the two phases of the survey. Since the two unique surveys shared a single source for their identifiers, which had duplications and variations in mapping, question phrasing, and an array of inconsistencies, this task would prove to be quite complex. The strategy to approach this would involve multiple phases on a limited set of the data, then scaling out the approach to the rest of the survey responses.

Step one would be to build out a table that ordered the survey question identifiers for the baseline survey, as well as a comparable table for the follow-up survey. From here, the next pipeline phase would match the unique question identifiers in a simple reference table. Following that step, we can filter for questions taking numerical response values and

Capstone Project for Masters in Data Science at University of Washington March 2020

pair them side by side as columns in a table housing the survey responses. Lastly, by differencing values in the paired columns, we can obtain values representative of changes over time for patients.

Multiple approaches were also employed in attempts to create predictively useful models. The first and simplest of these approaches was using forward selection to build multiple linear regressions.

The first step was capturing the correlation between each variable in the quality of life (QoL) section of the baseline survey and each other variable in the dataset. To accomplish this, the QoL variables were first divided into categorical and continuous variables. Many variables in the QoL section are not strictly categorical but are numeric discrete variables: multiple questions in the survey asked respondents to rate the severity of a symptom, record duration in number of hours, or report the number of instances of something. In these cases, it was determined that the variables would be appropriate outcome variables for regression modeling because the values do represent real numbers and the intervals between them have real meaning. Therefore, looking for a regression line of best fit does make logical sense. However, some sources suggest that treating discrete variables as continuous is problematic when the values can be near zero [8]. In this dataset, they can be near or even equal to zero; therefore we acknowledge this as a potential limitation of our approach and, for the purposes of checking correlation to other variables, some of these QoL outcomes were included in both the categorical and continuous groups. In this way, the results of different methods of measuring correlation could be compared. For categorical variables, a Chi-Square test of association was used. For continuous variables, both the R-squared of single linear regressions and a Pearson correlation were used.

Step two was the manual selection of outcome variables that appeared to have many strongly-correlated potential predictors. In this section, we discuss a model which aims to predict the likelihood that a patient will report feeling depressed.

Step three was the building of a model using forward selection, wherein predictors are added to a model one at a time, beginning with the most strongly correlated predictor, then proceeding to the next most strongly correlated, etc. In this step, before a predictor was added to a model, the survey

was checked manually to ensure that the method being used to measure correlation was appropriate for that variable. The response rate and distribution of the variable were also checked manually; if the apparently high correlation was an artifact of very low response rate (i.e., some columns that only contained data in one row appeared very strongly correlated to the outcome) then the variable was not added to the model. Variables were also skipped if they were too strongly interrelated to other variables already included in the model. For example, there are several questions throughout various sections of the survey which ask about depression or depressive symptoms. Including all of these variables would greatly increase the model's risk of multicollinearity.

The final predictive approach, the random forest, was chosen due to the model's ability to handle hundreds of predictor variables and determine feature importance via mean decrease in impurity. It also has built-in handling of null values. This approach was broader than the regression analysis, as all variables were considered and a different set of response variables were modelled. The response variables chosen were determined to be general indicators of 6 different categories of data: restlessness, feelings of depression, moods and worries, lifestyle, fatigue, and difficulty of concentration. All of them were viewed as categorical variables. Skewed responses were handled by using a log transformation.

For feature selection, low-fill features and outliers were removed and some summary statistics, such as number of medications, were created. Two feature sets were tested: one with all of the data and one with only summarized data and collapsed categories. As in the linear regression, some features were removed as they were very obviously correlated to the response and would not provide substantial analytical inference. An initial random forest was created to determine the 15 features with that decreased the weighted impurity the most. These features were used to test logistic regression, decision trees, and other random forests.

VI. RESULTS

The pipeline used for the longitudinal analysis was prototyped on the "General Quality of Life" and "Knowledge of Survivorship Care" survey sections. However, afterwards, the formatting differences throughout the rest of the survey would prevent the full dataset from completing a problem-free

Capstone Project for Masters in Data Science at University of Washington
March 2020

conversion. The “Reproductive Health” section proved to be a particular pain point here. Consequently, this portion of the analysis was limited to the aforementioned sections. For context, its key to know that differences were taken as baseline responses minus follow-up responses.

The results of the “General Quality of Life” longitudinal analysis imply that the majority of patients (60%) did not cite any difference in their general health at the time of each survey. 22% cited improvements, while 18% chose a lower response value in the follow-up than in their baseline. However, when asked how they would rate their health in comparison to a year ago from each survey, 41% chose the same value, 26% chose a higher value, and 32% selected a lower value. Figure 6 offers a visualization of the distribution of these response differences.

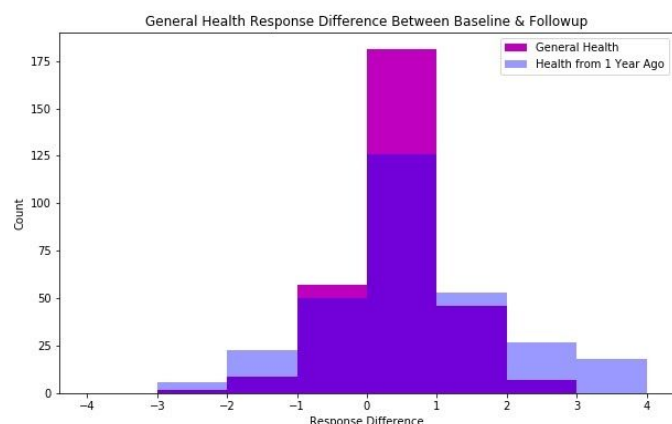


FIGURE 6.

In regards to how their health limited their ability to perform vigorous activities, 187 out of 303 (62%) respondents chose the same response in both surveys. 68 (22%) chose a higher response (more difficult), and 48 (16%) cited a lower response (less difficult). In all other activities, the number of responses with no change, at minimum, tripled the next most common response occurrence type, with only 19 total respondents citing a change in the ability to bathe or dress oneself, as opposed to 284 not changing their response

In the 4 weeks preceding each survey, respondents were asked how much time their physical and emotional health affected their regular daily activities. When how much either interfered with normal social activities, 55% kept consistent answers between both surveys, while 23% and

22%, respectively offered improved and declined responses. The effects for each of these causes individually follow in Table II and Table III.

Table II.

Physical health causes...		
<i>to cut down time spent on work or other activities</i>		
171 no change (56%)	72 increased (24%)	60 decreased (20%)
<i>to accomplish less than you would like</i>		
148 no change (49%)	77 increased (25%)	77 decreased (25%)
<i>limitations in in kind of work or other activities</i>		
142 no change (47%)	90 increased (30%)	70 decreased (23%)
<i>difficulty performing work or other activities</i>		
159 no change (53%)	71 increased (24%)	72 decreased (24%)

Table III.

Emotional health causes...		
<i>to cut down time spent on work or other activities</i>		
146 no change (49%)	73 increased (24%)	82 decreased (27%)
<i>to accomplish less than you would like</i>		
151 no change (50%)	72 increased (24%)	79 decreased (26%)
<i>doing work less carefully than usual</i>		
188 no change (62%)	52 increased (17%)	62 decreased (21%)

The results for the majority of responses in this section followed the general pattern seen thus far, with “no change” representing 50+% and the other two categories

**Capstone Project for Masters in Data Science at University of Washington
March 2020**

having a gap of no more than 2-7% between them at most. However, there were some that held results a little more extreme. Select cases are displayed in Table IV.

Table IV.

<i>Felt so down nothing could cheer you up in the last 4 weeks</i>		
211 no change (70%)	41 increased (14%)	50 decreased (17%)
<i>How often you felt worn out in the last 4 weeks</i>		
137 no change (45%)	86 increased (28%)	79 decreased (26%)
<i>How much of the time have you been happy in the last 4 weeks</i>		
199 no change (66%)	47 increased (16%)	55 decreased (18%)

The section on “Knowledge of Survivorship Care” exhibited more neutral distributions than the previous one. Here, patients answer questions about their level of confidence in their knowledge of their cancer and care plans. The leading question asks about the type of cancer the patient has. Relatively few respondents cited a decline (9%). 74% remained neutral, and 16% cited improvements. Numbers were comparable on the subject of the stage of cancer, at 11% declining, 20% improving, and 69% remaining unchanged.

Overall numbers appeared positive for this section of the survey. For various topics, patients tended to feel much more confident about their knowledge of cancership care by the time of the follow-up. The following table (Table V.) outlines some choice questions to demonstrate this point.

Table V.

<i>How often you should have testing to detect recurrent cancer</i>		
182 no change (61%)	83 increased (28%)	34 decreased (11%)
<i>Possible symptoms of recurrent cancer you should report</i>		

164 no change (55%)	99 increased (33%)	37 decreased (12%)
<i>Things you can do to prevent cancer recurrence</i>		
156 no change (53%)	114 increased (39%)	26 decreased (9%)
<i>Strategies for preventing long-term physical effects of cancer treatment</i>		
161 no change (54%)	107 increased (36%)	28 decreased (9%)
<i>The long-term emotional effects you may experience</i>		
156 no change (53%)	110 increased (37%)	30 decreased (10%)
<i>Which screening tests you should have to detect other cancers</i>		
144 no change (49%)	130 increased (44%)	22 decreased (7%)

A linear model (Figure 7.) was constructed that appeared to be successful in explaining a meaningful portion of the variance in patients’ likelihood of feeling depressed, but its predictive utility in a clinical setting is doubtful.

OLS Regression Results						
Dep. Variable:	feeling_depressed	R-squared:	0.608			
Model:	OLS	Adj. R-squared:	0.602			
Method:	Least Squares	F-statistic:	99.74			
Date:	Tue, 03 Mar 2020	Prob (F-statistic):	2.35e-75			
Time:	14:18:43	Log-Likelihood:	-384.33			
No. Observations:	393	AIC:	782.7			
Df Residuals:	386	BIC:	810.5			
Df Model:	6					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	4.8385	0.060	81.049	0.000	4.721	4.956
reduced_enjoyment	-0.2585	0.061	-4.266	0.000	-0.378	-0.139
being_a_burden	-0.1030	0.060	-1.724	0.086	-0.221	0.014
little_pleasure	-0.3194	0.049	-6.575	0.000	-0.415	-0.224
cant_control_worrying	-0.1549	0.047	-3.273	0.001	-0.248	-0.062
distant_from_others	-0.1181	0.038	-3.083	0.002	-0.193	-0.043
lack_of_sleep	-0.0954	0.035	-2.698	0.007	-0.165	-0.026
Omnibus:	16.894	Durbin-Watson:	2.053			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	22.327			
Skew:	-0.371	Prob(JB):	1.42e-05			
Kurtosis:	3.981	Cond. No.	7.58			

FIGURE 7.

The strongest predictors for likelihood of depression tended to be other mental health-related factors, including questions that are arguably asking about symptoms of anxiety. This is of dubious utility to medical professionals as the relationship between depression and anxiety is well

Capstone Project for Masters in Data Science at University of Washington March 2020

understood and established. The one potential exception is sleep; although depression and anxiety both have well-researched relationships to sleep, sleep is a physical and quantitative indicator that can be treated independently of mental health concerns. As this project continues and the pool of clinical data grows, it would be useful to examine prescription data to see if lack of sleep continues to be a useful predictor of depression when it is treated using drugs.

Perhaps the more interesting result of the attempt to build linear models that successfully predicted QoL variables is that some domains of potential predictor variables consistently failed to be correlated with any QoL outcomes. Most notably, pain-related predictors were not well correlated with any examined outcomes and were not suitable for use in any model.

The random forest models had varying degrees of success. Accuracies in predicting binary responses (low or high satisfaction categories) of restlessness, depression, mood, general lifestyle, fatigue, and concentration were 84, 84, 62, 75, 76, and 85 respectively. The logistic regression and decision tree models did not yield adequate accuracies. Prior to this, regression responses were tested but those models had very high mean squared error - likely a result of features only having answer columns of 1 to 5 and having few rows of patients after data cleaning. It was found that the detailed feature set did not increase the accuracy substantially over the reduced feature set.

Many models depended primarily on a small handful of highly important variables that were intuitively correlated with the indicator (Figure 8). Fatigue was connected highly to having little energy, feelings of depression were connected to anxiety, and general lifestyle with days of activity. As with the linear regression analysis, the utility of these findings are limited as many of these connections are well known. Unfortunately, random forest models do not retain much of the interpretability of other models, but individual trees can be inspected to gain some amount of insights into how the final model is built (Figure 9).

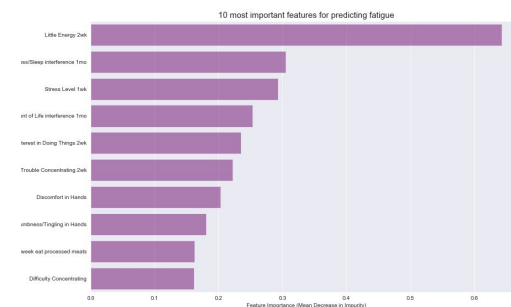


FIGURE 8.

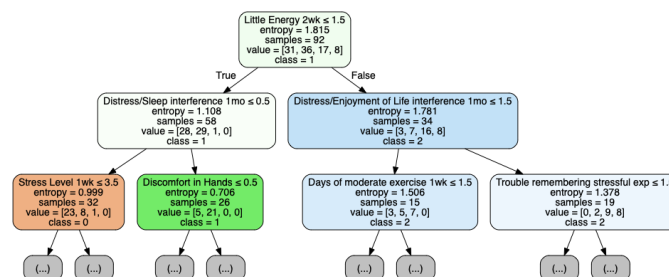


FIGURE 9.

Lastly, within the code files, a statistical breakdown is supplied on the clinical data for numerical fields containing at least 2 entries. The resulting table contains 16 descriptive fields over 600+ clinical features; far too large to represent in this paper. However, the interpretation of these results is also better left to someone holding more expertise in the field than this operative team currently holds.

VII. DISCUSSION/CONCLUSION

The biggest challenges present with the SIMS data are the knowledge barrier for the clinical data and the structure of the raw survey-related data. The significant amount of work that has gone into creating the data pipeline, data mappings, and reproducible analyses will be useful tools in generating more key insights from the SIMS data. That said, there are findings presented that yield valuable information about the survivorship program.

Gauging by the “General Quality of Life” and “Knowledge of Survivorship Care” sections, the patients are largely walking away from their treatment educated about their plans and personal health, and are not experiencing

Capstone Project for Masters in Data Science at University of Washington March 2020

notable detriments to their perceived quality of life. However, it should be noted that this is based on only limited longitudinal analysis of the patient responses. By branching this analysis out to other sections within the surveys, more evidence could still yield new information. However, preliminary findings during the discovery and exploration phases lean towards this finding more supportive evidence than not. Further adding confidence is that questions not listed in this paper predominantly shared the patterns found in those outlined here previously.

Novel findings from predictive analysis were limited as many uncovered correlations were well-established or obvious. Depression was related to mental health factors, and fatigue was related to sleep responses. Some domains, such as pain, were consistently found to not be strong predictors of quality of life. Random forest models were also found to predict some quality of life indicators accurately, but many of the features that contributed most to the models were intuitively correlated to the predictor in the same ways that were uncovered in regression analysis.

Several limitations of this data set are likely to be ameliorated by time and continued data gathering. We hope that our work provides a solid foundation for future analyses as the SIMS dataset continues to grow.

ACKNOWLEDGMENTS

This project was made possible through a collaboration with the Fred Hutchinson Cancer Research Center. The progress made in this project would not have been possible without the gracious assistance of Thomas H. Adams, Tabitha Headrick, Emily T. Silgard, Carly Strasser, PhD, Ben McGough, Justin M. Burge, Karen Syrjala, PhD, K. Scott Baker, MD, MSc, and Emily Jo Rajotte, MPH.

We would also like to extend sincere thank you to Capstone Director Megan Hazen for being our trusted guide and resource every step of the way.

REFERENCES

- [1] Fred Hutchinson Cancer Research Center. *Getting Active and Staying Active*.
<https://www.fredhutch.org/content/dam/www/research/patient-treatment-and-support/survivorship-program/survivorship-health-links/Getting%20Active.%20Staying%20Active.pdf>
- [2] Fred Hutchinson Cancer Research Center, Survivorship Clinic. *Health promotion through diet and physical activity*.
<https://www.fredhutch.org/content/dam/www/research/patient-treatment-and-support/survivorship-program/survivorship-health-links/Health%20Promotion%20Through%20Diet%20and%20Exercise.pdf>
- [3] Fred Hutchinson Cancer Research Center, Constance Engelking, Susan Leigh. *After Effects of Cancer Treatment*.
<https://www.fredhutch.org/content/dam/www/research/patient-treatment-and-support/survivorship-program/survivorship-health-links/After%20Effects%20of%20Cancer%20Treatment.pdf>
- [4] Fred Hutchinson Cancer Research Center/Livestrong, Hester Hill Schnipper. *Stress*.
<https://www.fredhutch.org/content/dam/www/research/patient-treatment-and-support/survivorship-program/survivorship-health-links/Stress.pdf>
- [5] Fred Hutchinson Cancer Research Center/Livestrong, Brad J. Zebrack. *Living with Uncertainty*.
<https://www.fredhutch.org/content/dam/www/research/patient-treatment-and-support/survivorship-program/survivorship-health-links/Living%20with%20Uncertainty.pdf>
- [6] Fred Hutchinson Cancer Research Center/Livestrong, Susan Leigh. *Body Image*.
<https://www.fredhutch.org/content/dam/www/research/patient-treatment-and-support/survivorship-program/survivorship-health-links/Body%20Image.pdf>
- [7] Fred Hutchinson Cancer Research Center/Livestrong, Katherine Walsh-Burke. *Sadness and Depression*.
<https://www.fredhutch.org/content/dam/www/research/patient-treatment-and-support/survivorship-program/survivorship-health-links/Sadness%20and%20Depression.pdf>
- [8] The Analysis Factor, Karen Grace-Martin, *When Can Count Data be Considered Continuous?*
<https://www.theanalysisfactor.com/count-data-considered-continuous/>