

Can Natural Language Processing Reveal Doctors' Attitudes toward Specific Medical Conditions?



H. Brooke Scoles
Keble College
Candidate no. 1027153
Word Count: 25,405

Submitted in partial completion of the
Master of Philosophy in Economics

Trinity 2019

Abstract

In recent decades, an increasing quantity of economics research has been conducted on biased beliefs and their impact on labor markets, education outcomes, housing and other economic activities. The discrimination that results due to biased beliefs is difficult to observe. Still, a combination of econometric techniques and inventive experimental design have provided convincing evidence that such bias exists. Research into biased beliefs and discrimination has mostly focused on gender and race, while very little has been done on disability. Within existing studies of disability there is almost nothing on "hidden" disabilities or disease, with no papers examining the influence of bias on patients. Systemic biases in the judgment of patients, if they exist, can potentially influence their care. Hence, biases are difficult to measure in a lab setting because this environment obscures behavior that is not easily detectable or that is intentionally hidden. One way to approach such an exploration is to look for evidence in language in anonymous settings. Natural language approaches, including LASSO-logistic regression and emotion dictionaries, provide the tools for this mode of analysis. The analysis was focused on a specific disease, ME/CFS, because of characteristics that make it a good candidate for researching attitudes and biased beliefs. Using this approach, it appears that medical decisions regarding treatment are not entirely objective and that they are influenced by incorrect beliefs. The language used by medical professionals shows that doctors' attitudes towards patients are not consistent but vary in line with different diseases. Such differences have economic implications, potentially lowering the quality of care, worsening health outcomes and lowering labor market productivity or participation.

Contents

1	Introduction	3
1.1	Introduction	3
1.2	Economic Implications	9
1.3	Alignment of Incentives	10
1.4	Economic Costs of Lost Medical R&D	10
2	Social and Political Motivation	13
2.1	Motivation for Focus on ME/CFS	13
2.2	Illustration of Data	24
2.3	Potential Influence of Gender	31
3	Literature Review	35
3.1	Literature Review	35
3.2	A Note on Discrimination	41
4	Data and Methods	45
4.1	Data Source Description	45
4.2	Data Collection and Processing	47
4.3	Methodology	50
4.4	Word Selection	57
5	Results	59
5.1	LASSO-Logistic Regression - Binomial Case	59
5.2	LASSO-Logistic Regression - Multinomial Case	61
5.3	Dictionary Correlations	68
5.4	Discussion and Implications	77

6 Conclusion	81
6.1 Conclusion	81
A Appendix	83
A.1 Least Absolute Shrinkage and Selection Operator (LASSO)	83
A.2 Cross-Validation	89
A.3 Logistic Regression	92
A.4 LASSO-Logistic Regression	95
Bibliography	97

1

Introduction

1.1 Introduction

In recent decades, bias and discrimination have increasingly become areas of economic research. These forces have had an impact on labor markets, education outcomes, and incarceration rates as well as other parts of the economy (Bertrand and Mullainathan, 2004; Bertrand et al., 2005). Discrimination is considered to exist when a minority group is treated less favorably than the majority group, despite both groups sharing identical characteristics. Discrimination is difficult to observe, but a combination of regression analysis, observational data and lab experiments have provided convincing evidence that it exists (Bertrand and Duflo, 2016). Sources of discrimination include preference-based bias due to correct beliefs, and bias due to incorrect beliefs (Bohren et al., 2018b). The latter case of incorrectly specified beliefs is considered to be inefficient in economics and to have negative welfare implications. This is the case that will be primarily considered in this paper.

A review of the economic literature on this topic located relatively abundant research on biased beliefs regarding gender and race, but very little regarding disability. Most of the existing research did not mention disability at all, including in Ondrich et al. (2003) or List (2004). Within disability, no papers were found on “hidden” disability, such as neurological disease. There was only one paper

found within a medical context, which looked at biases against referrals to female surgeons, yet there were no papers locatable looking at the influence of bias on patients. Systemic biases in judgment of patients, if they exist, can lead to incorrect stereotypes about that patient group and potentially influence their care.

Bordalo et al. (2016b) have formally modeled and tested stereotypes in a lab setting. While providing some useful information, the lab environment also presents limitations for measuring beliefs. First, not every bias is visible in behavior that is easy to observe. In addition, even if the behavior is easy to observe, it may not be detected in a lab if subjects are aware of expectations and take steps to hide their true beliefs. This would especially be a risk for biases that do not match social expectations (Bordalo et al., 2016a; Wu, 2017). In a medical context, doctors may feel pressure to appear impartial to improve credibility and trust or avoid lawsuits (Bartelme, 2005). They may also deny mistakes out of fear, with the possibility that their own errors have harmed or killed a patient (Nyquist, 2014). The way they speak in the lab may not be representative of their actual feelings about patients.

However, doctors do give evidence of their views in the language they use online. One way of approaching the problem of hiding true beliefs is to use natural language processing techniques on this anonymous data created outside of a lab. Gentzkow et al. (2017) makes the case that natural language techniques are particularly useful for detecting emotions and group beliefs in text.

Do medical professionals judge chronically ill patients differently depending on their specific disease? This has implications both for research priorities and clinically in terms of the efficiency and depth of the testing that a doctor should run to investigate an illness. If stigma against certain diseases is leading to underinvestment, then this implies that policy makers should reevaluate metrics used to make funding decisions. If diseases with similar severity face different attitudes from medical professionals, then a doctor may put more effort into treating a subgroup of patients even if their level of disability is similar. The patients who receive smaller effort in treatment are less likely to get better than patients who receive more effort. This

has economic impacts, including reduced human welfare from quality of life and mortality, decreased labor market participation and decreased productivity.

Beliefs have real economic implications. For example, work by Sarsons (2017b) has shown that beliefs about gender influences perceptions of workers' ability and that this impacts the labor market. Within a medical context, biases that result in lower quality of care can lower worker productivity or prevent them from achieving complete entry into the labor market. Sarsons (2017b) demonstrates that incorrect bias is a market inefficiency because it is not consistent with rational expectations. If expectations were rational, assumptions about the distribution of agent type throughout a population would not be distorted by incorrect beliefs. When applied to disability, this area breaks with the literature because disabilities can affect a person's abilities in ways that race and gender do not. One way to approach this issue is to compare disabilities with the characteristics they have in common, such as a similar impact on quality of life.

Language is not only important for revealing true beliefs, but also for determining behavior. Behavioral economics has shown that emotions impact upon decision-making and that one way of measuring emotions is through language. Bollen et al. (2011) measure public mood states using Twitter posts and show that it can predict movements in the stock market. Using meeting notes from central banks, their communication has been found to be a more relevant predictor of interest rates than contemporaneous policy rate decisions (Lucca and Trebbi, 2009; Bohren et al., 2018a). Chen (2013) has shown that language can influence a range of economic behaviors from savings rates, retirement allowances and health choices. Thus, it is possible that the language used by the medical community represents views that have real impact on patient care at the clinical level as well as research priorities at the policy level.

Given data about the language used by doctors, it is possible that studies using language to measure beliefs could be repeated in a medical context. I¹ thus hypothesize that the attitudes of doctors toward specific diseases can be captured

¹First person is used in the style of Deirdre McCloskey, author of *Economical Writing* (2000)

in the language they use to discuss them. I have hence designed and implemented a research strategy to test the hypothesis that natural language processing techniques on data from online discussions can reveal doctors' attitudes toward specific medical conditions. This is the paper's key contribution to the literature. Measuring these attitudes required gathering novel data from an online forum used by medical professionals, which appears to be the first time this data source has been used in economics research. A literature review did not reveal any other existing papers in the domain of economics using text or other data from this forum.

In order to focus the analysis, I direct most of the attention towards the disease Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). The reasons for this choice and supporting evidence are discussed in Section 2.1. I do not argue that it is the only stigmatized disease, or even the most stigmatized, but rather that it is a good candidate for analysis of any negative attitudes and stigma within the medical community. I have hypothesized that ME/CFS patients would face a high degree of bias relative to other diseases. If true, this would make it a good candidate for checking for effects by disease, being that the regression is most likely to show clear results for diseases with the biggest effects.

There are academics who make the case that there is currently under-investment in treating patients with ME/CFS (IOM, 2015a; Dimmock et al., 2016; Montoya, n.d.), which can also occur at the clinical level. Million dollar workups on patients with negative prognosis may not provide comparable benefits, but under-investigating cases can prevent the recovery of patients who may have been treatable (IOM, 2012; Burke, 2016). Attitudes, especially potentially stigmatizing ones, may play a role in these inefficiencies. Bias has been shown to be inconsistent with rational expectations and distort the market. Thus, if attitudes of medical providers depend on the disease type for patients with similar disability levels, this is a market inefficiency. Through analyzing text data from anonymous discussions by medical providers, I found evidence that doctors appear to be more judgmental when discussing ME/CFS relative to other diseases.

The discussion forum, Reddit Medicine, is intended as a place for doctors to share news, patient cases and medical anecdotes, as well as pose questions for discussion by other doctors. Users post anonymously, which means their speech may be edited less than the comments that they make in other places. Anonymity should allow for the data to capture what doctors genuinely believe. I have hence written a script to retrieve all the posts, comments, and sub-comments from the forum for 26 major diseases. Here, the text is inherently high-dimensional with thousands of variables. Processing steps were then needed to reduce the number of features in order to run regression, but methods for high-dimensional data were still required. Once in the database, I processed the text using these dimensionality reducing methods and created a word frequency matrix to summarize the data.

To select the relevant features from the remaining terms in the frequency matrix, I ran a LASSO-logistic regression. Here, I ran regressions for both the binomial case, comparing ME/CFS to all other diseases, and for the multinomial case, comparing ME/CFS to other diseases with similar characteristics. LASSO is a machine-learning tool used to choose relevant features from a dataset with many variables. The goal is to build a model that represents this data and determine which attributes best describe the response variable. This approach is helpful because, in text data, there are thousands of potential regressors, while it is not obvious which are important to include in the model. If each of the variables were added to the model as in Ordinary Least Squares (OLS), the model would be vastly overfitted.

Another benefit of LASSO is that it can determine the size of the coefficients for each variable as well as which subset of features are relevant when they are in a large number. It can do this even when the number of features is greater than the number of observations available (also called the $N > p$ case). Redundant attributes that do not add additional information are consequently removed by setting their coefficient to zero. The resulting model will have fewer features and will be easier to interpret (Fonti, 2017).

LASSO is similar to other optimization methods such as OLS, the difference being that LASSO adds a constraint on the absolute value of the parameters.

Since the absolute value is penalized, this requires standardization of the variables. Standardization has been done here. The size of the penalty parameter in the constraint, λ , is chosen through cross-validation - a method that splits the data into training and test sets and minimizes the error. The resulting $\hat{\beta}$ for each attribute will be biased, on average, but its variance will be smaller, leading to predictions that are more reliable (Wu, 2017).

In economics, LASSO has been used to predict movements in stock prices, changes in firm performance and returns to schooling (Belloni et al., 2011; Bandiera et al., 2017; Chinco et al., 2018). Wu (2017) uses LASSO to predict gender differences in language.

Using data from discussions on the medical forum, I use LASSO-logistic regression to test for differences in attitudes between conditions. If differences exist, these attitudes could influence medical decisions and treatment. The regression results show that medical opinion may not be objectively determined. Moreover, the comments here do distinguish between diseases, showing that doctors' attitudes are not consistent. Attitudes toward patients appear to be different according to the disease. This result holds when comparing diseases that have defining characteristics in common.

Next, I checked the correlation between the word frequencies for the diseases in the dataset and the emotions measured by the General Inquirer Dictionary, a dictionary developed for sentiment analysis at Harvard. As this dictionary does not have an emotion category for stigma, I created a custom dictionary to measure this and again checked the correlations with each of the diseases. I found that the stigma words were correlated with ME/CFS but not for the other diseases. In a blinded test, using five additional diseases, the new dictionary was correlated with one of the diseases hypothesized to face stigma from medical providers and with none of the diseases that were not. The full list of diseases is shown in Figure 4.2 on page 18. The presence of negative attitudes held by medical providers has several economic implications.

1.2 Economic Implications

Ill health has a substantial impact on human welfare in relation both to quality of life and life expectancy, but also through reduced utility gains from consuming goods and partaking in activities unrelated to health. It also diminishes an individual's ability to earn income. At the macro level, the economic consequences of disease include lower GDP and lower growth prospects. Improving the health of society has the economic consequence of directly increasing utility but also doing so indirectly, by increasing the enjoyment gained from consumption goods purchased with additional income.

ME/CFS is a useful disease to focus this topic, in part, because of its sizeable economic impact. Between 836,000 to 2.5 million Americans suffer from ME/CFS, while in the United Kingdom the number is estimated to be between 130,000 and 200,000. Meanwhile, 80% of patients are women (IOM, 2015b), while 25% of patients are severely ill, bedbound or housebound (Pendergrast et al., 2016; Valdez et al., 2019). In the UK, 4 in every 5 severe patients do not receive home visits by an NHS doctor and do not receive any medical care as a result (Action for M.E., 2015)².

Considering medical bills, care-taking expenses and lost labor market productivity, the disease costs the United States' economy approximately \$20 billion and the United Kingdom approximately £3.3 billion annually (IOM, 2015b; Optimum Health Clinic Foundation, 2017). The range of estimates is large, however, with a Centers For Disease Control (CDC) study estimating economic losses at \$1 billion just for the US state of Georgia (Lin et al., 2011). Lost household and labor force productivity makes up \$9.1 billion of the US loss. Stigma from health care providers may delay treatment and discourage investment in biomedical research. Without treatment, the impact of ME/CFS on each patient will be more severe. These patients may also have reduced participation in the labor market. For this reason, this paper attempts to measure these attitudes on the part of medical professionals.

²This may be an overestimate because Action For M.E. is an ME charity. However, it does not appear that this is tracked by any government agencies in the UK.

1.3 Alignment of Incentives

There is also an issue of misalignment of incentives. Treatment of some diseases may have a large social benefit, but a high cost to treatment - at least in the short term. In 2006, a Parliamentary All-Party Group inquiry into ME/CFS found that that:

There have been numerous cases where advisors to the DWP have also had consultancy roles in medical insurance companies. Particularly the company Unum Provident³. Given the vested interest private medical insurance companies have in ensuring CFS/ME remain classified as a psychosocial illness, there is blatant conflict of interest here.

Such a conflict would apply particularly to insurance companies whose policies have not included coverage for mental illness. An Unum internal report on ME/CFS stated that ME/CFS has been caused by "failure of coping mechanisms" combined with "entitlement philosophy" and an inability to keep up in an increasingly demanding economy. The report authors speculated that as "the American Dream is well out of reach for many," stress was increasing instances of ME/CFS which was made up of, "highly educated professionals slipping into self-imposed oblivion! [emphasis in original]" The firm noted that their "dollar exposure [to ME/CFS claims] is significant" and that they would "lose millions if we do not move quickly to address this increasing problem". The firm consequently resolved to present the disease as "neurosis with a new banner" and push treatments that would "increase [patients'] motivation to return to work" (Jackson, 1995). If this attempt at rebranding has been successful, it should be visible in the data within the terms medical professionals use to speak about the disease. When analyzing these attitudes, it is thus useful to keep in mind the misalignment of incentives between insurance companies and social welfare that may be contributing to incorrect bias.

1.4 Economic Costs of Lost Medical R&D

Language may signal decision-making that has real economic impacts in terms of the level of R&D devoted to medical research. If the attitudes of medical professionals

³Unum is a disability insurance company in the UK.

toward ME/CFS are more negative than toward other patients, one implication would be that medical R&D is less than the optimal level and that stigma is partly responsible for this underinvestment. In addition to contributing to diagnostic errors, this stigma may prevent policymakers from investing in research. If the level of investment is inefficiently low due to the presence of stigma, then the typical R&D framework does not account for this.

Grant and Buxton (2018) used the standard theory to measure the benefit of R&D. Every £1 spent on public research in the UK was associated with an additional £.99 of private expenditure on R&D and a 15%-18% return on investment in terms of the overall impact on GDP. Linking each estimate to a specific medical intervention, these studies first estimated the amount of government dollars spent on a given disease, then estimated the amount of health gains that could be attributed to those investments, then found the net monetary benefits: cost of delivering treatments - value of health gained ⁴ (Grant and Buxton, 2018).

This model is:

$$\Delta y_{i,t} = \Gamma \Delta y_{i,t-1} + \alpha(\beta y_{i,t-1} + \mu) + \gamma + \epsilon_{i,t} \quad (1.1)$$

where $\Delta y_{i,t} = [public_{i,t}, private_{i,t}, sales_{i,t}]$

Here, $y_{i,t}$ is a 3x1 vector where i is the disease and t is the year. Γ is a matrix of autoregression coefficients. α is a vector of correction coefficients adjusting for the departures from the cointegration (long-run equilibrium) equation. μ is the vector of intercepts in the cointegration equation and γ is a vector of drift terms. The coefficients reveal how public sector investment, private sector R&D and pharmaceutical sales respond to deviations from long-run equilibrium, revealing a cointegration relationship where public and private sector expenses complimented each other. The elasticity of private R&D with respect to public research was nearly

⁴Each quality-adjusted life year (QALY) was given a value of £25,000 which reflects the opportunity cost to the NHS, which is how much it usually costs the NHS to generate an extra QALY. QALYs are a measure of health that combines length and quality of life. One QALY is equivalent to 1 year of life in perfect health. A QALY of .8 would represent someone who is 20% disabled.

1. Therefore, each additional £1.00 of public biomedical research would result in an increase in private pharmaceutical R&D of nearly the same sum (Sussex et al., 2016).

Any attempt to measure the impact of government spending on health research needs to account for the spillover effects of that spending on private sector R&D. Within this context, spillover effects are the impacts of investments by one organization that benefit other organizations, whether public or private. For example, when pharmaceutical companies use information created by public sector research to enhance their drug development process to bring new medicines or technologies to market. Spillovers can also happen the other way around, thus entailing the additional need to account for the spillover effect on private sector R&D. Thus, estimated elasticity, the social rate of return and private R&D investment may be combined to estimate the total GDP gain (Grant and Buxton, 2018).

This standard investment theory cannot completely account for underinvestment in medical research. If there are systematic biases held by the medical community, these biases would contribute to underinvestment and the true benefit of R&D would be higher than measured by this framework.

My HIV patients, for the most part, are healthy and hearty thanks to three decades of intense and excellent research and billions of dollars invested. Many of my CFS patients, on the other hand, are terribly ill and unable to work. I split my clinical time between the two illnesses, and I can tell you if I had to choose between the two illnesses, I would rather have HIV. But CFS, which impacts a million people in the United States alone, has had a small fraction of the research dollars directed towards it.

Dr. Nancy Klimas, AIDS and CFS researcher and clinician, University of Miami

2

Social and Political Motivation

2.1 Motivation for Focus on ME/CFS

This section will focus on the social and political motivation for using ME/CFS to study incorrect bias. Section 3.1 will cover the theoretical background in economics as well as the precedent in the economic literature for using the chosen methods. In that section, I will also refer to the body of knowledge in economics for studying stereotyping and bias. In this section, however, I will discuss the societal factors motivating the study of medical attitudes and the political environment that makes ME/CFS an illuminating example.

The decision to focus the analysis specifically on the neuroimmune disease ME/CFS was done made for a combination of reasons. Diseases with a large degree of stigma will have the largest signal in the data. I hypothesize that conditions like ME/CFS, which are more frequently represented in the media and have a complicated political history, have more stigma than diseases that are not talked about as often or are not considered controversial, such as Multiple Sclerosis. Due to the severity of the illness, if there is a negative attitude from medical providers, then there is a larger gap between the existing attitudes and the attitudes that would be most helpful for treatment.

Patients with less severe diseases may not need as much medical care or attention from doctors. For illnesses that do not need as much care, negative attitudes are still undesirable. However, for these conditions, there is a smaller gap between the negative attitudes and the level of sympathy or concern required for proper medical treatment. Diseases with larger gaps between their severity and the quality of treatment from medical professionals face a greater degree of incorrect bias. ME/CFS is a good disease for studying incorrect biases because it has both these qualities: an involved political background as well as a severe impact on quality of life.

The New York Department of Public Health (NYDPH) in the United States describes Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) as "a serious, debilitating, chronic disease that affects multiple body systems, including the nervous system, the immune system, and the body's production of energy." The department explains that, "symptoms of ME result from an abnormal response by the immune system, most often to an infection. This can impact the brain and other systems of the body" (NYSDOPH, 2018). The Center for Disease Control (CDC) describes it as a long-term "disabling and complex illness" that "affects many body systems" (CDC, 2018b).

As early as 1990, studies have shown the patient's immune systems have reduced natural killer cell function which impairs their ability to attack pathogens (Klimas et al., 1990; Nguyen et al., 2017). The disease also alters multiple mitochondrial functions and reduces the capacity of patient metabolism (Naviaux et al., 2017). A damaged metabolism diminishes the body's ability to convert food into energy, build proteins, and eliminate waste (Alberts et al., 2002). In every measure of energy capacity, the cells of ME/CFS patients are impaired relative to healthy controls (Naviaux et al., 2017). Combined with brain and spinal cord inflammation, this condition has led some doctors to hypothesize that ME/CFS is an autoimmune disease (Nakatomi et al., 2014; Montoya et al., 2017; Davis, 2017). Meanwhile, research on patient metabolism has found metabolites in ME/CFS patients up to 16 standard deviations out of range compared to healthy controls (Davis, 2016). Figure 2.1 shows the scope of this difference.

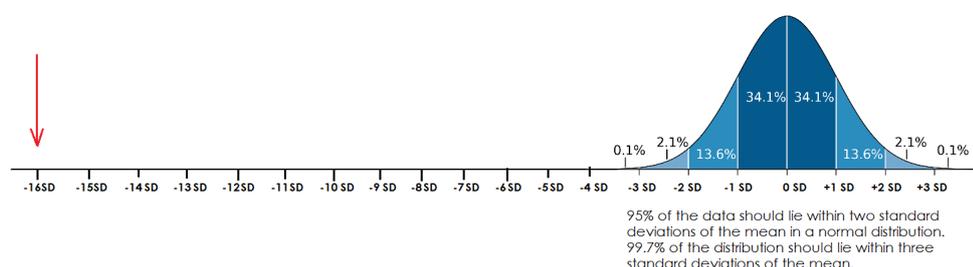


Figure 2.1: Visual Description of 16 standard deviations (Rally, 2012)

In 2015, an Institute of Medicine report reviewed thousands of medical publications to conclude that ME/CFS was “an acquired, chronic multisystemic disease biological in nature”, including symptoms such as “immune, neurological and cognitive impairment”. The report then “stresses that this is a medical – not a psychiatric or psychological – illness” (IOM, 2015b).

Ron Davis, the Director of the Stanford Genome Technology Center, has called it “one of the most urgent areas in medicine today” (Wright, 2018). In 2013, an article in the Atlantic described Davis as “one of the greatest innovators currently working,” then going on to write that “a substantial number of the major genetic advances of the past 20 years can be traced back to Davis in some way” (Allan, 2013). He has been recognized for developing the method that made it possible to complete the Human Genome Project (Science, 2001).¹ He has been studying the disease since his son Whitney Dafoe became ill. His son is now bedridden with severe ME and cannot speak or eat, being fed through a PICC line.² Regarding the severity of ME/CFS, Ron Davis has said:

This is a much more serious disease than many of the other things that people are worried about. It’s more common than MS, it’s more common than Parkinson’s disease, it’s more common than AIDS. This is probably the last major disease that we know so little about. And it’s because of its nature that it’s been hidden. The severe patients are often

¹James D. Watson, PhD, who helped establish the Human Genome Project, works with Davis on the End ME/CFS Project as a member of the scientific advisory board (OMF, 2019).

Watson won a Nobel Prize in 1962 for his work on DNA structure (Foundation, 2019).

²PICC stands for a peripherally inserted central catheter, as a tube inserted into a vein or, as in Whitney’s case, directly into the heart and used for IV medication or nutrition.

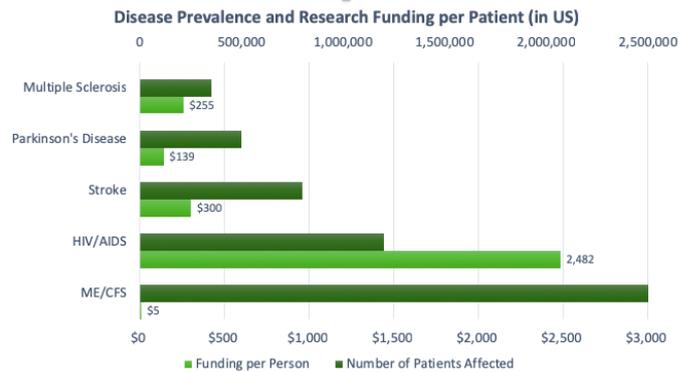


Figure 2.2: In the United States, ME/CFS affects more people than many major diseases, but receives lower funding per patient. Funding is \$5 per patient per year.

just in their home being looked after by someone and no one knows they exist. But it can get very severe, people have tried to make some measures in terms of debilitating illness, it's generally viewed as worse than many other diseases that have been ranked in terms of quality of life. - Ron Davis, PhD Director of the Stanford Genome Technology Center (Wright, 2018)

Ron Davis is not the first person to make this observation. The advocate and director of the documentary film 'Unrest', Jen Brea, reasons that it is specifically because ME/CFS is so severe that medical professionals have come to overlook it. In an interview in 2017 she said, "Sometimes people think, 'If it's really this bad, how could it have been ignored?' Because people with this disease have such a hard time leaving their home or engaging in advocacy. When you can see us, we look normal, and when we're sick, you're not going to see us. It's created this problem that no one can see" (Brea, 2017).

In addition to its severity, ME/CFS has appeared suitable for studying biases of medical providers because of the tone in some of the comments by medical providers online. The quantitative analysis starting in Section 4 can help determine whether negative comments are outliers or trends. The qualitative examples in Section 2.2 help illustrate the harshness in tone that has motivated further investigation. For instance, Medscape, a website for doctors, has covered the IOM report and received 304 comments, the majority of which were dismissive toward patients³

³Investigative journalists working on ME/CFS noticed the general attitude of the Medscape comments and described their tone in The New York Times article cited above.

David Kaufman, MD California

I will assume that the first two comments are by physicians. Let's begin with a suggestion: find a doc who treats these patients and spend one week in their office. You will be astounded at the degree of illness and misery (and I would hope a little shamed). You will hear patients describe how they cannot wash their hair, they cannot stand for more than 30 seconds, that writing an email leaves them exhausted, and on and on. I am talking about people that were PhDs, software engineers, venture capitalists, major real estate developers, entrepreneurs, you name it. These are not lazy people looking for a disability check. Several times a week I hear the same comment: I wish I had AIDS or Leukemia.

Regarding biomarkers of disease, it is correct that there is no single test comparable to an HIV pcr or an A1C for diabetes. But virtually every patient has profoundly low natural killer cell function, about 80 % have undetectable vasopressin levels, about 95% have mutations of the MTHFR gene (compared to 30% in the general population), and nearly every patient has documented abnormalities of the HPA axis. The real problem is our abysmal lack of understanding: why do these patients get sick, what is the cause, what starts the spiral? Perhaps research dollars might solve this problem.

But I would submit that not knowing the CAUSE of disease does not mean there is NO disease. We treat hypertension and don't know the cause. We treat MS and don't know the cause.

Figure 2.3: Dr. Kaufman, MD defends the IOM report in the comments section of the New York Times.

(Tucker, 2015; Rehmeyer, 2015).

Doctors specializing in ME/CFS treatment, such as former Medical Director of the Open Medicine Institute and HIV/AIDS researcher Dr. Kaufman MD, have publicly pointed out the juxtaposition of the stigma with the severity of the illness. Kaufman's comments about this contrast are in Figure 2.3. In these remarks, he argues that increasing funding for research would improve medical knowledge (Rehmeyer, 2015). These observations spotlight an important relationship: the presence of stigma, or bias, creates an inefficiency in investment (for a discussion, see Sarsons, 2017b; Bordalo et al., 2016b) by limiting the willingness of clinicians and policymakers to direct resources toward a disease. However, this effect also runs in the other direction; that is, if there is a lack of biological evidence for a condition due to lack of research, this may further perpetuate skepticism by medical providers and stigma. As previously shown, evidence of physical abnormalities is not lacking in

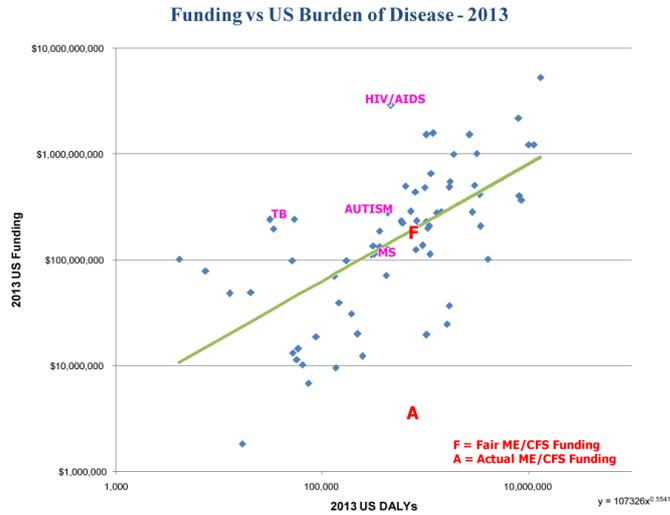


Figure 2.4: NIH funding versus burden of disease (US).

this patient group (Klimas et al., 1990; Nakatomi et al., 2014; IOM, 2015b; Wilshire et al., 2018). However, there is lack of understanding among medical providers about how to test for these abnormalities and treat them (Blease et al., 2017). The limits of medical education are also described through the examples given in Section 2.2.

While ME/CFS can be severe, it has not been a funding priority. In the United States, the National Institute of Health (NIH) spends less money overall per year than it spends on male pattern baldness or hay-fever. As shown in Figure 2.2, there is currently less than \$5 spent on ME/CFS per patient annually in the United States. In economic theory, the benefits from funding depend on the marginal benefit of an additional dollar of research. Consequently, measuring the marginal benefit requires a full accounting of factors including scientific opportunity. Short of this, metrics such as the present level of funding or disease burden shed light on what the marginal benefits might be. Diseases with low funding levels may have more low hanging fruits remaining in research. Furthermore, diseases that are very severe stand to gain more in quality of life for each extra dollar than diseases that are mild. This is also true for diseases affecting large numbers of people.

Disease burden is one of the central factors the NIH uses to determine which diseases to fund. This factor is considered alongside the incidence of scientific opportunity and research interest. The NIH has combined data from several studies

estimating the burden of major diseases to compare how the agency's funding matches this need (NIH, 2016). Dimmock et al. (2016) have noticed that while the NIH had combined this data for many diseases, this has not been done for ME/CFS. They have used the World Health Organization's Disability Adjusted Life Years (DALY) measure combined with the number of people impacted to calculate this burden of disease for ME/CFS.

When comparing the estimate to the NIH analysis of research funding versus DALY, they found that federal funding is less than what would be expected given the disease burden. While this work was limited by the availability of data, their rough estimates found that to match the level of other diseases, annual funding would need to be 25 times higher (Dimmock et al., 2016).

Figure 2.4 shows that funding is also low relative to the impact on quality of life compared to other major diseases. Each additional dollar of research would impact upon a significant number of patients and have the potential to improve quality of life. Estimating the marginal benefit of an additional dollar of spending is outside the scope of this research, but initial indicators such as the number of people impacted, disease severity, economic cost and present level of saturation in the field indicates that the marginal benefit to additional research funding may be significant. Hence, this is an area that merits further study.

The marginal benefits of research also depend on the scientific opportunity in the field. If a study impacts many people but investment is not likely to be fruitful, then it is not necessarily inefficient to invest resources elsewhere. However, researchers working in the field are optimistic and several ME/CFS researchers believe, due to the low level of investment, there are many potential discoveries that would be easy to obtain given more funding (Montoya, n.d.; Tucker, 2019). Dr. Jose Montoya, Professor of Infectious Disease at Stanford University, has said "We think this is solvable within a lifetime. If this disease is given the money that is given to other diseases, we will have the solution within a 5-year period" (Montoya, n.d.). Systemic biases in the medical community could otherwise lead to an underestimation of the level of scientific opportunity.

Language that implies opacity, such as "complex" or "vague," does not imply outright negative stereotypes the same way as words like "lazy" or "malinger," but it can still have an impact on investment. Francis Collins, NIH Director, has alluded to the impact of this tone, saying that the air of mysteriousness surrounding ME/CFS has been unhelpful to the research environment. In an interview with *Science*, the director said that many researchers were put off from researching problems in this area, thinking that "Maybe this is an unsolvable problem, let's just work on something else" (Cohen, 2015).

Determining whether a problem is solvable or not contains an element of subjectivity. If the discourse surrounding a condition emphasizes uncertainty and complexity, then it could undermine the willingness of researchers to take on ME/CFS projects. The language used to discuss a disease could influence whether or not it is to be funded beyond the objective marginal benefits and costs of doing so. While there are scholars arguing the scientific opportunity in the field is high (see Montoya (n.d.); Tucker (2019)), the tone of medical providers is what this paper intends to measure.

The authors of the 2015 Institute of Medicine report subsequently noted that:

...the committee was struck by the relative paucity of research on ME/CFS conducted to date in many areas related to this disorder. Remarkably little research funding has been made available to study the etiology,⁴ pathophysiology,⁵ and effective treatment of this disease, especially given the number of people affected.

Another factor that potentially contributes to negative attitudes regarding ME/CFS is psychosomatic explanations for the illness (Geraghty and Esmail, 2016). Despite the advances in biomedical understanding on the part of ME/CFS researchers over recent decades, psychosomatic explanations for the disease have prevailed in the early history of ME/CFS,. Proponents of psychosomatic theories have argued that patients harbored 'unhelpful illness beliefs' that they have a physical illness that is limiting their abilities. To help patients recover, therapists

⁴cause or causes

⁵the disordered processes of the body and its systems associated with disease

must alter these beliefs so patients can overcome their fear of exerting themselves (Blease et al., 2017; Rehmeyer and Tuller, 2017). The vestiges of this initial history still impact upon policy priorities, diverting funding away from biomedical research and leading to adverse patient outcomes (Geraghty and Esmail, 2016; Wright, 2018).

To some extent, acknowledging the effect of psychological factors may appear reasonable. The biopsychosocial framework, often cited in connection with ME/CFS, asserts that all illnesses are a result of a combination of biological, psychological, and social factors (Geraghty and Esmail, 2016). Such an observation is not particularly controversial, as most would acknowledge a combination of physical and psychological factors contribute to any disease process. However, a criticism of the biopsychosocial model is, that more often than not, it is used to emphasize psychological causes over biological ones (Butler et al., 2004; Ghaemi, 2010, 2011).

One weakness of psychosomatic theories is that they are not falsifiable (Popper, 1989).⁶ According to these theories, there are no physical signs of disease. Fatigue, pain, cognitive deficits are a result of ‘sickness behaviors.’ However, as an additional part of the theory, any measurable biological abnormalities eventually discovered are considered to be a consequence of psychological illness, not a cause (Geraghty and Esmail, 2016; Maes and Twisk, 2010). This contradiction then prevents falsifiability.

In addition to being impossible to falsify, psychosomatic theories have been associated with adverse patient outcomes. The biopsychosocial model discourages practitioners from conducting extensive testing on patients with ME/CFS, arguing that further testing perpetuates the illness and drains resources (Geraghty and Esmail, 2016). However, minimizing medical investigations comes with risks, such as wrongly diagnosing and under diagnosing patients. A study of specialized chronic fatigue clinics in the UK found that 43% had an alternative missed diagnosis and 37% of patients were referred incorrectly (Devasahayam et al., 2012).

The idea that patients’ beliefs perpetuate the illness prevents them from accessing social care. The conclusion of theories that propose patients hold ‘dysfunctional

⁶Falsifiability as a philosophical theory was introduced by Karl Popper in 1989. Statements must be falsifiable to be considered scientific. Theories that are not falsifiable are then classified as pseudoscience.

illness beliefs' is that they are adopting the 'sick role.' Therefore, symptoms can be reversed through the patients' efforts and if symptoms are not reversed, it is because the patient was not working hard enough (Blease et al., 2017; Geraghty and Blease, 2018). This framework, particularly the assumption that patients can change their mindset and recover, has been used by the Department of Work and Pensions (DWP) in the UK to deny disability benefits (Shakespeare et al., 2017). Such an approach cuts the total budget required for disability programs, as its supporters contend that disability payments will "make patients worse" and "foster a culture of dependency" (Faulkner, 2016).⁷

Psychosomatic theories have also been used to justify lack of investment in research and treatment. For example, from the limited government funding available for studying ME/CFS, this psychological model has diverted millions of pounds away from investigating the biological abnormalities in patients into psychological research (Wright, 2018). UK Psychiatrists used these funds to develop the treatments of Graded Exercise Therapy (GET) and Cognitive Behavioural Therapy (CBT), which have assumed that the patients could use exercise and positive thinking to return to their previous levels of health. This research has since been discredited (Vink and Vink-Niese, 2018; Wilshire et al., 2017).

Studies promoting GET and CBT do not rely on any objective measures of health. In one influential case, reanalysis of the data found no statistically significant difference between treatment groups. It was later found that the authors had weakened their outcome measures to such an extent that patients could get worse throughout the duration of the study and still ultimately be classified as recovered (Wilshire et al., 2018; Rehmeyer and Tuller, 2017). The studies and resulting treatments in question here were debated in UK Parliament twice where they were referred to as a "medical scandal" (of Commons, 2018, 2019).

⁷The author of this report stated that while he expected this to be two weeks of work, this led to three years of research as well as the "grim realization that the problems in this area were so serious and widespread that that no-one could hope to explore them comprehensively" (Faulkner, 2016).

In addition to being ineffective, this treatment model encourages overexertion which can lead to long term damage (Geraghty and Esmail, 2016; Blease et al., 2017; Geraghty and Blease, 2018). In the United States, a 2013 Institute of Medicine Report found that “exertion of any sort can adversely affect several organ systems and many aspects of patients’ lives, often seriously and for long periods” (IOM, 2015b). Overexertion has been found to cause cell damage, while therapies such as CBT that teach patients to ignore symptoms encourage activity that can cause their condition to decline (Twisk and Maes, 2009; Kindlon and Me, 2011; Carruthers et al., 2011; Vink and Vink-Niese, 2018). As a result, US health agencies, including the NY Department of Public Health and the Center for Disease Control, have removed their recommendations for CBT and GET.

Although these treatments are discredited, this research has not yet trickled down into clinical practice. If patients refuse treatment due to fears that it will worsen their condition, some psychiatrists advocate that patients – including children – need to be admitted to enforced rehabilitation (usually in a psychiatric unit), citing treatment noncompliance or enforcing attendance in school (Dr. Nigel Speight, 2014). Blame for treatment failure is therefore shifted from the medical professionals onto the patient or their family.

In the UK, the National Institute for Health and Care Excellence (NICE) guidelines have recommended that children with ME take part in education. However, this has been interpreted by child protection authorities as a physical presence in school instead of access to remote education in the home. Tymes Trust, a UK charity, has found that hundreds of parents have been referred to social services, investigated for child abuse and been threatened with having their child taken into government care (Burgess, 2018; Colby, 2014). As this is an ME charity, the figure may be an overestimate, but there are currently no official figures to compare.

However, the issue is starting to be recognized by policymakers. In Parliament in March of 2019, MP Carol Monaghan raised the issue with PM Theresa May, stating that "ME/CFS affects approximately 25,000 children in the UK, when the parents of these children make the extremely difficult decision to remove their children

from a program of Graded Exercise Therapy, many of these families have child protection proceedings triggered against them. In the midst of this Brexit chaos, will the PM commit to looking into this issue?" (Monaghan, 2019).

This social and medical context makes the neuroimmune disease ME/CFS a condition with the potential to reveal stigmatizing attitudes. It is likely to carry a greater amount of stigma than other diseases such as depression due to its political background. Diseases that are discussed with larger degrees of animosity will be more easily measured using the regression analysis. Due to the severity of ME/CFS, it can be compared with other illnesses that significantly impact upon quality of life, such as Multiple Sclerosis or Lupus. It may be possible to check if it is discussed with tone that conveys less sympathy or urgency than these similarly severe conditions. The next section thus gives anecdotal examples of hostility from medical providers that motivated this research into whether this could be detected systemically in the discussion.

2.2 Illustration of Data

The following section includes qualitative examples of the effect that the paper attempts to measure. The following examples are comments by doctors posting on new articles about ME/CFS, or posts about specific patient cases where most of the discussion on the thread concerns ME/CFS. These are all from medical professionals, rather than professionals from other disciplines or "laypeople" who may be commenting on the same forum.

Many of these comments appear to be written by medical professionals from the United States. The name Myalgic Encephalomyelitis (ME) is used by the World Health Organization (WHO) and used as a diagnosis throughout Europe, while Chronic Fatigue Syndrome (CFS) is the official name in the United States. Choosing the name CFS has drawn criticism for defining the disease by a symptom and appearing to trivialize its impact, similar to calling Tuberculosis the Chronic Coughing Disease or calling Alzheimer's Chronic Forgetfulness Disease. The focus on fatigue is also nonspecific and leads to confusion with a wide range of diseases that

include fatigue as a symptom, such as Multiple Sclerosis, Leukemia, or Depression. However, CFS is the current official name in the United States and the name used by doctors and researchers working there (Rowe et al., 2017).

Chronic Lyme, adrenal Fatigue, CFS, Ehlers-Danlos⁸, Fibromyalgia⁹, SIBO¹⁰ and if you really get a believer, Morgellons.¹¹ They get shuffled around every few years. - Nurse Practitioner (NP)

They likely found no change because ME/CFS is almost certainly a form of somatized depression/anxiety and not a B-cell-mediated disease. Giving these people immune suppressants seems a very expensive/dangerous placebo. - Third-year Medical Student (M3)

Yeah, we should combine that diagnosis with fibromyalgia and call it "subacute nothingness." - Radiology

In this case, and to a lesser extent with fibromyalgia, the problem with the PR¹² about this disease and the problem with physicians due to perceived "hostility due to physician lack of knowledge and empathy" is the *fucking patients themselves*. - MD Emergency Medicine

Look, I work in hippyland. I treat people people [sic] who think homeopathy and crystals have medicinal value. I treat people who think they have chronic candida and are allergic to scents and wifi (and the last is not hyperbole). I have no problem with people who have weird beliefs about their illness. But every single person who has come to me with CFS (or CFIDS whatever), drives me 'round the bend with a gazillion demands for tests and forms so they can get every possible accommodation in the known universe. The fact that almost every CFS patient is an upper middle-class white woman makes those demands even more ridiculous. I have patients in wheelchairs after CVAs¹³ who don't ask as much.

The person who thinks they are allergic to scents might want to have a school or work accommodation to avoid the scents they dislike, but the CFS patient who walks into my office with a Starbucks coffee wants a handicapped plate (permanent natch'), 80 hours of IHSS to do their laundry and clean their place, discounted electric bills, permanent SSDI,

⁸Ehlers-Danlos Syndrome (EDS) is a group of conditions that affect connective tissue.

⁹Fibromyalgia Syndrome (FMS), is a condition that causes long-term, widespread pain in the body among many other symptoms.

¹⁰Small Intestinal Bacterial Overgrowth

¹¹MedicineNet describes Morgellons disease as "a delusional disorder that leads to the belief that one has parasites or foreign material moving in, or coming out of, the skin"(Roxanne Dryden-Edwards, MD, n.d.).

¹²Public Relations

¹³Cerebrovascular Accident, also called a stroke

and to pay \$200/month for their section 8 housing by getting me to write a letter saying they spend an inordinate amount of their SSDI check on medically necessary bullshit to treat their 'illness'.

And twice I have had patients like this try to get me fired or go to the state board because I refused to acquiesce to their demands. Maybe if they weren't malignantly entitled crazies, they wouldn't need the PR benefit of a name change. – MD Emergency Medicine

Ladies and gentleman [sic], we've reached a point in society now that lay people can make up diseases and the medical establishment has to accept them. Read that article, it's just oozing of "why were my feelings hurt?!" The internet has done countless great things for our society, but it also allows the masses to dictate discussion, whether it's informed or not. I simply can't wait for my mandatory Morgellon's CME¹⁴, can you guys? - Nurse Practitioner

this response is very well said. You described all of my thoughts about CFS (and FMS¹⁵ to a certain extent) very eloquently. I'm just a pre-med now, but I am totally and completely fascinated by the subset of syndromes characterized by chronic fatigue and/or chronic pain, but with no known physiological/pathological origins (i.e., I am not talking about things like ankylosing spondylitis, clear orthopedic deformities, etc.).

I have worked as a scribe in the ED for a while, and without a doubt, your socio-economic observations about people with these diseases are spot-on: they are almost invariably (but not always) upper-middle class white or middle-class white women. Often college educated. It is interesting to examine their co-morbidities, which often include depression, anxiety, lower back pain, etc. I am interested in learning more about why these patients are suffering (or why they PERCEIVE themselves to be suffering) and maybe doing research on chronic pain in medical school. I'm especially interested in finding ways to help chronic pain patients re-establish healthy, productive lives in ways that don't involve narcotic use. It is sad to simply see them in the ER¹⁶ over and over again. It is even sadder to see the narcotic/benzo/other controlled substance addictions that develop... - 4th-Year Medical Student

I still find this one of the hardest illnesses to take seriously. I regularly work with fibromyalgia, chronic pain, ptsd, and depression, and yet when I hear about CFS I have to suppress a reflexive scoff. I don't know what it is about it that makes it so hard to accept, and I hope some progress is made soon. - Canadian GP, Specialist in Poverty & Addictions

¹⁴Continuing Medical Education

¹⁵Fibromyalgia Syndrome

¹⁶Emergency Room

...as Dafoe's condition got worse, he stopped replying in sentences, and began answering text messages with just a 'Y' or an 'N'. Then those, too, stopped coming. Dafoe, now 34 years old, can no longer speak. He communicates with his parents through small motions, such as ripping holes in the shape of hearts in paper towels. I took that from the article. I can't help but be skeptical. Moreover, it's with said father starting an IV in a home setting for treatment of the disease. It just seems too dramatic to not be at least in some part mere drama. - Infectious Disease and Critical Care Medicine¹⁷

I refuse to believe that 1/200 pts¹⁸ have some sort of chronic post-viral disease state that causes symptoms that mirror depression.

I do not shut off the possibility that some sort of post-viral disease state exists in a much smaller population. I unequivocally refute that it would be so incredibly common and feel that many patients with clinical depression are pushing for a non-psychiatric illness diagnoses. I honestly feel I could create "16s disease" populate it with vague symptoms, get some loose associated physical findings and be just as scientifically valid as CFS. - Nurse Practitioner

Easily my least favorite patients to encounter. I had someone with CFS tell me last week that 'doing the dishes' should be considered exercise, that doing so depletes all her energy and then subsequently she needs to sleep for 14 hours. Towards the end of a totally unproductive and long-winded visit full of tears, she of course perks up and hands some disability forms to my attention (as is tradition). - Rural GP

Appears to be a politically phrased consensus statement to me. Systemic exertion intolerance¹⁹ is simply a nice way of saying 'I can't or won't get off my ass'. Symptoms can comprise a syndrome, but that does not make it a disease with objective, consistently reproducible findings. - MD Neurology

CFS is such BS. My mom worked 50-60 hour weeks with a hgb of 5²⁰ and never complained about it much except that she got a bit dizzy when gardening one of her days off and decided to get that checked out.

There is a huge percentage of the population that needs to, in the words of my med school rheum attending, "sack up" instead of hiding behind their diagnoses. - Internal Medicine

¹⁷This doctor is referring to the description of Whitney Dafoe in an article linked on the forum and dismissing it as "too dramatic" to be true.

¹⁸patients

¹⁹This was a proposed alternative name for ME/CFS that takes account of physical activity worsening symptoms.

²⁰Hemoglobin (Hb or Hgb) count. A hemoglobin count of 5 is considered extremely low. Low is defined as less than 13.5 grams of hemoglobin per deciliter.

My GP colleagues see a lot of these patients on a very regular basis...normal bloods, normal clinical investigations, negative rheumatology/cardiac/Neuro/endocrine workup. Also, severe polypharmacy issues and problematic lifestyle factors in the majority.

Can find the energy when they want to complain or get their meds early though – Specialized Pharmacist (Clinical)

This. Because for some reason people can't accept somatic symptoms as being somatic.

The next time disability paperwork comes in front of me for "fatigue" and your cardiac workup is unremarkable, I'm tossing it in the trash. This is getting ridiculous.

I speak as a cardiology NP who is tired of seeing age 30-40 year-olds claiming they just don't feel like working anymore. – NP Cardiology

there is often secondary gain associated with these chronic illness conditions. - 2nd-Year Medical Student (MS2)

And the ****overwhelming**** majority of patients who have complaints and you can't find a single thing wrong? They're almost all faking it/trying to get opioids. – Surgical Resident

Well, be aware that patients who come to the ER to have their chronic medical problems treated are probably the wackos. And I've seen plenty of them, too, in 30 years of family medicine practice - people who wallow in suffering. – Emergency Medicine

It's possible that the quotes above are outliers and do not reflect the general sentiment held by medical providers. This is the purpose of running the LASSO-logistic regression. For example, the following three quotes show a more sympathetic sentiment than from the doctors above.

I'm not going to fight you on this, but I would encourage you to be a bit less cynical. Just because you haven't experienced the same challenges your patients have, doesn't mean you should discount their complaints as invalid. Plus, often your patients can tell when you're annoyed or uncomfortable with the reason they're coming in for. In my mind, it's unprofessional. – Genetic Counselor

As someone who is often quick (and probably too quick) to dismiss things like CFS or Fibromyalgia, I appreciate your comment and think you make an important point. In a profession that prides itself on evidence-based practice, we have to make sure we don't close ourselves off to new evidence that may contradict current practice.

That aside, I'm jazzed about the gut microbiome²¹ and its implications on our health. I honestly believe that it will be one of the biggest sources of breakthroughs we will see in my lifetime. – 4th Year Medical Student

CFS causes cancer, heart disease and dementia, just so you're informed. Also, I was unaware that a serious debilitating illness had to be immediately terminal to warrant research attention. Should we stop researching ways to treat lupus and MS? Additionally, the majority of CFS patients are predominantly bedridden with severe cognitive dysfunction, rendered entirely disabled and often sicker than cancer patients and late-stage AIDS patients. Quality of life and functionality is WELL below those of patients with other common autoimmune and neurodegenerative disorders. – RA Cognitive Neuroscience

It's very hard to wrap my mind around something that just doesn't have any reproducible laboratory or imaging or functional (i.e. EMG, EEG, etc.) findings with our available test methods. But it's also hard to dismiss these people who are active, high achieving, functional people who get sick and then just don't get better. I really do wonder about the possibility of a bunch of different etiologies. And some of them probably *are* functional/conversion/psychiatric, which certainly must make studying the whole thing more difficult. – Pathology Fellow Year 5 (PGY5)

In these comments, it can be seen that medical professionals hold three primary misconceptions about the illness. First, their comments do not show an awareness of selection bias. They mention that most of their patients with suspected ME/CFS are upper-middle class white women, but the patients seen by any given doctor are not necessarily representative of the patient population as a whole. ME/CFS can impact any age and socioeconomic group. A 1999 study by Jason et al. found that minorities and people with lower levels of education had the highest prevalence of the disease (Jason et al., 2017).

Second, small amounts of exertion by an ME/CFS patient, such as brushing one's teeth or showering, can cause a worsening of symptoms (CDC, 2018a). In these comments, the doctors use the patient's level of disability and resulting lack of movement as evidence that they are not trying to improve.

²¹Gut microbiota are the bacteria/organisms that live in the digestive tract. They are a key part of the body's immune system and have been implicated in ME/CFS.

Lastly, it is implied throughout the threads that the patients are seeking secondary gain in the form of disability benefits or accommodations. In the case of ME/CFS, there is physiological evidence available if doctors run the correct tests. From this sample of comments, it appears that many doctors do not seem to know this. As mentioned above, there are tests for metabolites that show severe ME/CFS patients as much as 16 standard deviations away from healthy controls (Davis, 2016). When other tests are returned as normal, the patient is assumed to be malingering. There is no acknowledgement of the possibility of having requested the wrong labs.

ME/CFS isn't the only disease stigmatized on these forums. One post asks, "Is it common for certain conditions/diagnoses to be blacklisted at hospitals? What can be done when this happens?" Several responses from doctors mentioned that Ehlers-Danlos Syndrome (EDS) patients have effectively been banned, making comments such as "[EDS patients] take not just a lot of resources, but a lot of emotional bandwidth and can be hard to please. They're not usually officially blacklisted, but usually the response is: 'we don't have anyone here who's an expert in EDS.'"

Ehlers-Danlos that people are very fond of having these days is the "hypermobility type." – Pathology

Saying patients are "fond" of being ill presumes that being unhealthy is a choice or a desired outcome. It implies that there is secondary gain to being sick. The presumption of secondary gain assumes a desire to be sick rather than a desire to be well on behalf of the patient. The former blames the patient for their illness. Patients with diseases such as cancer, who are not accused of seeking secondary gain, may be treated with more sympathy by providers and given more thorough care.

The comments about EDS show that ME/CFS may not be unique in the way it is spoken about online. It is possible that doctors speak this way about everyone, or that the comments that are negative are not representative of the way that most people write on the forum. This is the motivating purpose for using LASSO.

The purpose of this paper is not only to see what words are used to describe common diseases, but also if there are any differences between disease groups. If certain types of language are common in comments across many disease types, then

they will also not show as predictive of a given disease category. LASSO will choose words as predictive when they have a systematic association with that disease.

2.3 Potential Influence of Gender

It is possible that there is also a gender element to this type of language. Dusenbery (2017) has researched gender bias in medicine and found that most conditions described as controversial in the media share characteristics in common with ME/CFS insofar as they are underfunded, under-researched and predominantly impact upon women. Natural language processing techniques on media commentary would augment this line of study.

At the same time, this is not to say that men are not affected. Dismissing physical symptoms as psychosomatic can and does happen to men as well. For instance, 20% of ME/CFS patients are men and they make up approximately 400,000 ME/CFS patients in the United States alone. Here, I argue only that there is a gender component to the disbelief patients face and that this is relevant to understanding both the political history and stigma around the disease.

In the book *Doing Harm*, Dusenbery (2017) , it was revealed that many foundational studies by the National Institute of Health (NIH) in the US enrolled only men; while the law changed in 1993, those studies have not been repeated including women. Due to their varying hormonal cycles, women were thought to be more complicated to research. However, Dusenbery (2017) maintains that the reason women were excluded from research is the same reason that it is meaningful to study them; that is, there may be differences in their biology.

On aggregate, women are now represented in human trials, but they are still underrepresented in key research areas such as heart disease, cancer and HIV. When women are included, the results are not usually analyzed by gender. Moreover, the requirement to include women applies only to human trials. Most preclinical trials only involve male lab rats. As a result, drugs with the potential only to help women do not make it to human studies. These issues are compounded by another problem: when diseases that specifically impact women are studied -such

as autoimmune diseases, gynecological disorders and chronic pain conditions - they receive meager research funding compared to diseases of similar severity that affect mostly men. This creates a situation where even well-meaning doctors do not have the tools to treat women (Dusenbery, 2017).

Even in solidly established diseases, a gender bias exists. Women later diagnosed with strokes, autoimmune diseases, endometriosis or brain tumours are often initially diagnosed as being chronic complainers or attention-seeking or "just tired" (Newman-Toker et al., 2014; Kennedy, 2016; Dusenbery, 2017).

Heart attacks, for example, are a condition mostly thought of as impacting men. Yet young women are twice as likely to die from heart attacks than men from the same age group (Mosca Lori et al., 2005). They are also seven times more likely than men to be sent home from an emergency room during a heart attack, while these discharges are associated with increased mortality (Pope et al., 2000). More women than men die from heart disease each year, but not even 20% of GPs are aware of this (Mosca Lori et al., 2005). The traditional signs of cardiac arrest were developed from studies that only included men. However, even when narrowing down samples to women who present with chest pain, they still experience diagnostic delays (Lichtman Judith H. et al., 2015). A 2007 study found that women and men were equally likely to be diagnosed with heart attacks in hypothetical examples, except when stress was added as a symptom. Then, only 15% percent of doctors diagnosed correctly in women while 56% did in men. The presence of anxiety led doctors to believe that women's symptoms were psychosomatic, while male equivalents were more likely interpreted to be physical even when anxiety was present (Chiaramonte, 2007).

In a qualitative study from Yale, women who recognized their symptoms as being likely due to a heart attack delayed going to the emergency room owing to fear of being labeled hypochondriacs. This fear appears to be justified (Lichtman Judith H. et al., 2015). Some doctors will be open about their explicit training. Dr. Adam Splaver, a cardiologist, admitted, "In training, we were taught to be on the lookout for hysterical females who come to the emergency room" (Archive, 2019).

This may explain why, in addition not to considering women's diseases to be in immediate need of funding or research, the medical community sometimes does not consider women's diseases to be "real" at all. While "hysteria" is no longer used in the medical literature, descriptions such as somatization, conversion disorder, psychosomatic and functional disorders have taken its place. The most frequently used is medically unexplained symptoms or (MUPS or MUS). While this may not seem immediately alarming, medically unexplained symptoms is often used to mean medically unexplainable. Hence, 70% of patients with medically unexplained symptoms are women (Dusenbery, 2017).

As Dusenbery points out, women are more likely to have medically unexplained symptoms because government agencies have put little effort into explaining them. Each time a woman sees a doctor about unexplained symptoms, regardless of whether these symptoms are later diagnosed, it supports the idea that women complain of symptoms without an organic cause. There is currently no system for feedback to the original doctor when it turns out a disease was misdiagnosed. Thus, clinicians have an inaccurate view of their own error rates²² (Berner and Graber, 2008). This creates a self-perpetuating cycle where women present with unexplained symptoms and doctors are quick to assume psychological causes and stop searching. In these cases, it becomes unlikely doctors will find evidence contradicting their initial assumptions.

As a result of this cycle, women find themselves in a double bind. Doctors do not understand their diseases because there is not enough research, while they do not trust women's own account of their symptoms which, in turn, further discourages research into those conditions. Women are more likely to have diseases that are initially assumed to be psychosomatic: Fibromyalgia, Vulvodynia, Migranes, Interstitial Cystitis, Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and chemical intolerance. These conditions are poorly understood because little research funding has been put into understanding them (Dusenbery, 2017). In addition to

²²This is also seen in accounts in the ME/CFS community where patients progress to being housebound or bedbound. Their doctors often assume the patient stopped coming because they got better when in reality, they were too ill to return.

studying comments by medical professionals, as this paper has strived to do, it would be worth comparing the language used in news articles. It is possible that most of the diseases described as controversial in the media are diseases that are underfunded, under-researched and mostly impact women.

The economic impacts of women's health are significant. By extension, this significance also applies to any stereotypes or incorrect beliefs held by medical providers that influence women's health. Women make up half the population while over 70% of adult women participate in the formal labor market in the UK compared to 78% of men (Office for National Statistics, 2019). In the US, the numbers are 57% of women compared to 69% of men (Bureau of Labor Statistics, 2018). They represent just under half of the labor force in the UK and US, 46.5% and 45.8% in 2018. Women's illnesses affect productivity and have significant economic impact. The labor market impacts of chronic disease include reduced employment, earlier retirement and lower income. The impact of health on social expenditures is also significant. In the EU, an average of 1.7% of GDP is set aside for disability payments and paid sick leave annually, which is larger than the cost of unemployment benefits. Chronic disease causes the early death of half a million people across the EU and the loss of 3.4 million productive life years (Union, 2016).

Thus, the fact that ME/CFS patients are 80% women, combined with the problems women face accessing treatment, make ME/CFS a useful disease to use for examining stigma and incorrect bias. The methods by which I approach this are described in the next section.

3

Literature Review

3.1 Literature Review

In recent years, digital text data has become more available for use in social science because an increasing amount of human interaction and communication has moved online (Gentzkow et al., 2017). Unlike traditional types of structured data such as prices, interest rates or callback rates, text is unstructured and includes news articles, social media posts, central bank meeting notes and political speeches (Hansen, 2018a). This data can complement traditional types of structured data and has been used increasingly in economics in a variety of applications. It has been shown to be useful in explaining outcomes of economic importance (Bohren et al., 2018a; Gentzkow et al., 2017). While text can be analyzed in a qualitative manner, here many economists have been analyzing text data quantitatively (Hansen, 2018a). Similar to traditional data sources, unstructured text data can be used for descriptive, predictive or causal analysis.

For example, text from financial news outlets and social media has been used to predict movements in price (Antweiler and Frank, 2004; Tetlock, 2007). Engelberg and Parsons (2011) have gone as far as to separate the causal effect of news on stock prices from other sources of correlation by exploiting the behavior of investors with access to different media sources during the same event (Engelberg and Parsons,

2011). In macroeconomics, text from Google searches has been used to forecast changes in inflation and unemployment. Scott and Varian (2015) has narrowed down common categories of trends and used variable selection methods to determine which categories are associated with consumer confidence and key economic indicators.

Gentzkow and Shapiro (2010) have used text to determine political slant in newspapers by comparing the articles to congressional speeches. They have expanded this analysis to determine the economic forces, such as consumer demand or political party of the news outlet owner, that would contribute to this slant (Gentzkow and Shapiro, 2010).

Most social science studies of stereotyping analyze objectively via measurable discrimination, such as callback rates for applicants or initial offers for products (Ondrich et al., 2003; List, 2004; Bertrand et al., 2005). However, bias can also be detected in language. Bohren et al. (2018a) has found that the sentiment of replies on a common mathematics Q&A forum varying according to the gender of the username of the poster. Furthermore, behavioral economics has repeatedly shown that behavior is affected by emotions. Bollen et al. (2011) found that the mood of society, as measured by Twitter posts, affects collective decision making. Using time series analysis, they found that moods after elections and holidays were predictive of closing values on the Dow Jones Industrial Average (DJIA). These findings contradict the Efficient Market Hypothesis (EMH) which states that stock markets follow a random walk and cannot be predicted with accuracy above 50% (Bollen et al., 2011). Instead, they show that economic and commercial indicators can be predicted to some extent and that indicators can be found in social media such as blogs or Twitter posts. Similarly, I argue that sentiment of the professional medical community has the potential to distort health markets, which could lead to inefficient treatment paths and underinvestment in biomedical research. The first step here is to argue that medical professionals are not always objective in their perception of disease. There are differences in perception that depend on the disease, as argued in this paper.

Text is inherently high-dimensional. It is possible to analyze an economics text using the same techniques that are used for high-dimensional data in other fields. In *Text as Data*, Gentzkow et al. (2017) explains that if there are w words in each document \mathcal{D}_i and a vocabulary of p total words, the documents have a unique representation of dimension p^w . For instance, 30 word Twitter posts using only the most common 1000 English words would have 1000^{30} dimensions. In order to make analysis possible, the dimensionality of the data thus needs to be reduced, which is the case even when using methods for reducing high-dimensional data such as LASSO or Latent Dirichlet Allocation (LDA) (Blei, 2003; Gentzkow et al., 2017).

LASSO and LDA are both supervised machine learning techniques used for variable selection and prediction. Supervised machine learning is also used to tackle classifications problems where labeled data is available.¹

LASSO, the method used in this paper, stands for Least Absolute Shrinkage and Selection Operator and is a penalized regression method. It is used to select the most relevant features relating to a response variable, especially in cases where there are thousands of potential regressors. It has been used in linguistics to narrow down the words most likely to predict demographic characteristics (Eisenstein et al., 2011). It has also been used to analyze the text of Congressional speeches that identify the most partisan words and phrases (Gentzkow et al., 2016).

LASSO is also being used in economics to analyze stock price movements, returns to schooling and firm performance (Belloni et al., 2011; Bandiera et al., 2017; Chinco et al., 2018). It could be that LASSO performs well out of sample because it is picking up on statistical artifacts, but it turns out that the predictors identified by LASSO are often economically meaningful. Chinco et al. (2018) finds that despite LASSO using statistical rules to identify predictors and not economic intuition, the variables that it identifies as meaningful are economically significant. In forecasting stock prices, LASSO has most often identified news about economic fundamentals as having the greatest predictive power.

¹This is in contrast to unsupervised learning, where the data only contains independent variables and where the goal is to determine the relationship between them.

In Wu's paper, *Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum*, text data has been used to look for differences in sentiment by gender. Using comments from Economics Job Market Rumors (EJMR), the author looked for evidence of differences in the way economists spoke about young women entering the field in comparison to men. The author scraped 250,000 threads over six years from the EJMR Forum. Posts were screened for gender classifiers, including "he," "she," and "him," which were then assigned to a gender category. Each post was added to a matrix categorizing the frequencies of each word that appeared. The most common 10,000 words from the comments were chosen for analysis. Using a LASSO-logistic regression, a subset of features was selected as having the greatest predictive power for whether a given comment was about a male or female.

The words with the greatest predictive power for comments about female economists included "hotter," "attractive," "tits" and "horny." while the words with the greatest predictive power for comments about male economists included "philosopher," "keen," "textbook," and "motivated" (Wu, 2017). One criticism of this technique is that breaking text down into individual words loses context. For example, if a word or phrase has the word "not" in front of it, the text actually has the opposite meaning from that implied by the word on its own. However, the EJMR forum is a forum where economists post about the economics job market and discuss the research and potential of top economists as well as their peers. If the word in a given comment is "hotter" or "attractive," even if that comment really says "not hotter" or "not attractive," it can be inferred that the poster is not discussing a given economist's work.

Similarly, on forums discussing the care of patients, one might expect that doctors would be discussing common treatments or strategies, as well as frustrations related to managing particularly complex diseases. One difference from the EJMR case is that this study uses comments from a forum where doctors often go to vent frustrations; hence, criticism of context must be more carefully dealt with. Reddit Medicine, unlike EJMR, is not a professional forum. One may argue that some comments made in a casual context are not acceptable in the workplace. In the

same way, comments on a forum where doctors often vent or complain may be acceptable so long as they do not affect their work.

However, in order to claim that beliefs expressed in casual forums do not impact patient care, one would need to argue that biases, including unconscious ones, are kept separate from work even in stressful circumstances. Bertrand et al. (2005) document that implicit beliefs can be measured and that they influence behavior meaningfully. They explain that situations that involve time pressure or stress accelerate mental processes lend themselves to implicit discrimination in an attempt to reduce "cognitive load," (Chugh, 2004; Bertrand et al., 2005). Furthermore, behavior influenced by implicit attitudes is more likely to occur in situations that are ambiguous and where many other non-discriminatory explanations for the behavior exist (Bertrand et al., 2005; Bohren et al., 2018b). Time pressure, stress and ambiguity are all likely to be present in rushed medical practices where there could be many other explanations for any given treatment decision. Lastly, if it is the case that doctors can abandon prejudices expressed in casual forums outside of their workplaces, it is still possible and relevant to use this approach to detect differences between diseases. If all diseases were spoken of with a hostile tone it would show only potential imprudence in expression, which may even be understandable after doctors have dealt with an emotionally charged patient situation. However, differences between diseases in tone would show that there is an additional relative bias against ME/CFS patients beyond what is normally found when discussing other diseases.

Many comments include words such as lazy, crazy, wacko and hysterical. Similar to the gender example, if a medical professional is using the word lazy to describe a patient or group of patients, it is relatively safe to say they are not discussing treatment or how to help them. Even if the comment says, "not crazy," it is clear the discussion is about mood states or mental capacity rather than about treatment. This is clear without knowing the history or politics behind the specific condition, which is not the case for all words, however.

Other comments include words like "somatized" or "psychosomatic," used to express the manifestation of psychological distress such as physical symptoms, and

"controversial" or "skeptical" which could have ambiguous meaning depending on the context. In the context of extreme anxiety, a doctor might say, "these chest pains are psychosomatic," which might not imply stigma. However, if the word was used to dismiss the patient experience, to escape responsibility for treating them, to blame the patient, or provide an explanation for not investigating further, the word could be assumed to be negative or perpetuating stigma. Within the context of this study, words are considered to have a stigmatizing tone if they imply suspicion, disbelief and doubt. The methods used to categorize these words will be described in Section 4.

Treatments based on psychosomatic theories may seem harmless. However, patients who do not receive necessary biomedical treatment may be prevented from recovering precisely because their conditions are misdiagnosed as psychosomatic. Severe patients who would have otherwise been working will be out of the labor market, not paying taxes, potentially receiving state benefits and taking from the system instead of contributing to it, thus relying on care from friends and relatives. The labor and funds required for caretaking then take resources from other market activities caretakers would otherwise be participating in.²

One search of the literature did not reveal any research on the economic costs of misdiagnosing patients' illnesses as psychosomatic. However, a 2016 IOM report *Improving Diagnosis in Health Care*, revealed that each year 12 million patients in the United States experience some form of diagnostic error. As a consequence, one out of every five cancers was misdiagnosed leading to delayed or incorrect treatment. Most Americans will receive at least one diagnostic error in their lifetime with 1/3 of these cases being life threatening. The report estimates that diagnostic errors cost the United States over \$100 billion/year (Balogh et al., 2015). Diagnostic errors lead to 80,000 preventable deaths per year and reduce the physician morale and limit public trust in the system (Hayward, 2002).

Heyhoe et al. (2016), Quality and Safety Researchers at the Bradford Institute for Health Research in the UK and the University of York, examined the role of

²Disability advocates would emphasize that people's lives have value beyond what they can contribute to the market. For the purposes of an economics dissertation, I am addressing mostly the economic costs here.

emotions in patient safety and found that emotions such as anxiety, anger, disgust and horror influence clinical decision making. Working with hostile or friendly peers affects both individual judgments as well as collective ones. They also looked at emotion as a trait and give evidence that doctors with a tendency toward a 'negative' affect influenced patient safety negatively in turn. These impacts included reduced accuracy and cognitive performance, as well as increased incidence of patient falls, hospital acquired infections, and medication errors. Moreover, the effect of emotions is higher when there is risk or uncertainty involved (Heyhoe et al., 2016). Iedema et al. (2009) establish a similar result. Their data also show that the affect of medical providers can be changed through shared narratives³ and that this can improve collective attentive to patient safety.

3.2 A Note on Discrimination

In the economics literature, discrimination has a very specific meaning (Ewens et al., 2014; Bertrand and Duflo, 2016). In particular, the gender literature in economics often measures a difference in compensation relative to an agent's type. If there are differences by gender in compensation for agents with the same ability, then this can be said to measure discrimination. This also holds for other signals such as the number of referrals for female surgeons of the same ability as male surgeons, or the number of positive votes on a forum for questions posted by female-named accounts versus male-named accounts (Bohren et al., 2018b). These studies assume that genders are of equal ability, or generate data and randomize across gender, or find some way of measuring the agent's true type (Sarsons and Xu, 2015; Sarsons, 2017b,a; Bohren et al., 2018a). The same applies to studies of race. When the outcome variable is not grouped by gender or race, but by disease, then this analysis becomes more difficult.

First, the literature on stereotypes focuses on beliefs about an agent's abilities relative to their true abilities. Then, outcomes like compensation can be measured relative to those predicted by the agent's true ability. In the case of disability,

³Narratives defined as the way medical professionals collectively discuss experiences.

the agent's type is their level of illness and the outcome is how much access they possess to public or private medical care.

Second, it is difficult to measure the type of the agent when the type is the severity of their condition. Having knowledge of how ill patients are is relevant to measuring whether medical discrimination is preference-based, biased due to correct beliefs, or biased due to incorrect beliefs. Discrimination due to preferences or correct beliefs has an unclear impact on welfare, although the literature has decided that discrimination due to biased incorrect beliefs is inefficient⁴ (Bohren et al., 2018b).

The preferences of medical professionals should not play a role in treatment decision-making. However, the case of discrimination in accordance with correct beliefs still needs to be ruled out. With the forum data used here, there is no way of measuring the type of individual patients discussed. Without measurement of individual characteristics it is not possible to precisely quantify discrimination. However, it is possible to find proxies for the 'type' of the patients at the aggregate level. Showing that the true average type of ME/CFS patients is not one of "faking," "malingering," or "lazy" – in essence showing that ME/CFS patients on average are acting neither more entitled or aggressive, nor comprise an unduly large number of healthy individuals merely claiming to suffer from the condition - could be the subject of its own dissertation. To do this, type would need to be broken into two dimensions: severity of illness and patient personality characteristics. Severity of illness could be proxied for using measures of average quality of life and the level of physical abnormalities in ME/CFS patients relative to healthy controls. These measures have been used in studies elsewhere to rigorously establish disease severity for ME/CFS patients (Klimas et al., 1990; Nakatomi et al., 2014; IOM, 2015b; Pendergrast et al., 2016; Nguyen et al., 2017; Naviaux et al., 2017; Montoya et al., 2017; Davis, 2017; Wilshire et al., 2017; Vink and Vink-Niese, 2018; Wilshire et al., 2018; Valdez et al., 2019).

⁴Discrimination due to common knowledge of correct beliefs is called statistical discrimination and is not considered inefficient. Discrimination due to preferences is called taste-based discrimination and is not desirable according to economic theory (Bertrand and Dufló, 2016).

Patient personality type and behavioral disposition could be measured using psychological tools that measure agreeableness, but can also be reasonably assumed to be consistent across diseases with similar characteristics. Without the individual data, there is no proof that the ME/CFS patient group do not in fact have the above characteristics, but it is sensible to assume they would not differ substantially from patients suffering from similar diseases, affecting similar demographics. If data shows that ME/CFS patients suffer attitudes generally negative at least in comparison with patients of similar conditions, that is evidence of biased group perception among medical professionals.

Jason et al. (2011) and Jason et al. (2017) show that MS, Lupus and ME/CFS share many characteristics in common. MS and Lupus are similar to ME/CFS in severity. They also are studied as part of the same category of disease as ME/CFS and thus share many of the same symptoms; both are autoimmune diseases, which ME/CFS is suspected to be (Davis, 2017). MS and Lupus also impact a similar demographic in that they mostly impact women (Jason et al., 2011, 2017). Even if ME/CFS patients behaved differently from other people and negative perception reflected that, we have little reason to expect they would be different from MS or Lupus patients.

Thus, I take as given throughout this paper that ME/CFS patients type in terms of personality is not less agreeable than the type for patients with other major diseases and the severity of their diseases is comparable as well. Patients with comparable characteristics such as people with MS and Lupus should be broadly similar. A comparison that reveals differences in tone should reveal underlying biases. Since ME/CFS patients should be on average the same type, I argue that differences should be interpreted as discrimination.

Furthermore, personality traits such as aggression or hostility are also a function of how people are treated by those around them. In the discrimination literature this is known as self-fulfilling prophecy (Glover et al., 2017). Self-fulfilling prophecies are perpetuated by confirmation biases and occur when treating a minority group in line with stereotypes about that group results in a shift in the underlying

group distribution. Eventually, the population will shift and the distribution of type will match the initial incorrect bias. In regards to gender, this happens if women are stereotyped as being bad at math and pushed out of math courses, the resulting distribution of skill in the population will be that women truly do have lower mathematical abilities than men (Bordalo et al., 2016b; Sarsons, 2017b; Bohren et al., 2018b).

Personality and behavior, as an agent type, could be a function of the way patients are treated by the medical profession and become a self-fulfilling prophecy in a similar way. I assume in this paper that even if ME/CFS patients exhibit different personality and behavioral characteristics from other patients of similar conditions, this is more likely a function of how they are treated by medical professionals than any inherent difference between the groups. I assume this, because there is no reason for ME/CFS patients to be of significantly different personality type than MS or Lupus patients, for example. If the data supports a comparative conclusion to the effect that ME/CFS patients are subjected to more negative attitudes than patients with other similar diseases, then it provides evidence for medical bias based on incorrect beliefs.

4

Data and Methods

4.1 Data Source Description

I collected the data for this study from the forum Reddit Medicine, an online forum that is part of the larger site, Reddit.com. Reddit is a news aggregation and discussion site made up of a collection of forums organized by topic called "subreddits." Subreddit names start with /r/ which is also part of the url where the forum is located. They cover a wide range of subjects and include /r/music, /r/food, /r/photography and /r/science. There are several health and medicine-related subreddits. The data used for this analysis comes from /r/medicine, which is also called Meddit (Nicol, 2018).

Users share news or content, including links, text posts and images, while they comment on posts shared by other users. The content submitted is then voted on by members of the community. Posts that have collected the most up-votes appear in the top of the subreddit and, eventually, if they are popular enough then they will appear on the Reddit front page (Franco, 2017).

Reddit is an American site and 53.9% of visitors are from the United States, but people can post from all over the world. Following the US, 8.2% of visitors were from the UK and 6.3% from Canada. In March 2019, Reddit ranked as the 6th most visited website in the United States and 21st most visited worldwide with

542 million monthly visitors, 234 million of those visitors being unique (Nicol, 2018; Internet, 2019).

Each subreddit has its own rules and moderating team. Administrators have set strict rules against harassment and spent considerable resources on moderating the site (Ohlheiser, 2017). For */r/medicine*, the rules include restrictions like the following:

- All submissions must be made by users who have set "flair," which indicates their medical background.
- Users cannot post personal medical situations.
- Users cannot post any identifying or protected medical information.
- Users may not ask for medical advice from the doctors on the forum.
- Posts from users whose post history indicates an agenda will be locked, preventing other community members from commenting or interacting with the post. For example, if a user mostly posts in Diabetes subreddits, then they may not post about Diabetes on Reddit Medicine.
- If a user posts on the forum who has comments or posts in their history that contradict the posts on the forum, especially with regard to medical background, they will have their post locked. If this continues, then that user will be banned from the forum (*r/medicine*, 2019).

To imitate a medical professional on this site, a user would need to have a separate account where they only post about medical concerns and must never contradict themselves. This is a barrier to entry for non-medical professionals, ensuring that most of the posts on the forum with user flair indicated are from actual doctors, nurses, medical students and other medical professionals. The opinions of doctors on this forum may not generalize as those who post may be different in type from doctors more broadly.

Posts tend to be skewed toward medical students and residents who are younger, rather than older working doctors. However, because the sample of doctors is possibly not representative of all doctors, this does not mean the data cannot be informative. A sample of students could have particular value. Students have

the most current medical education, entailing that the posts in this data are representative of what is currently being taught.

4.2 Data Collection and Processing

The diseases are taken from the study, *The Health-Related Quality of Life for Patients with Myalgic Encephalomyelitis / Chronic Fatigue Syndrome (ME/CFS)*, which compared ME/CFS to 20 medical conditions by quality of life. Results and diseases list are shown in Figure 4.1 on page 48. They have replicated the results from other studies and found ME/CFS patients have the lowest Health-Related QOL of the 21 conditions measured, worse than patients who have suffered stroke, cancers and multiple sclerosis (Komaroff et al., 1996; Falk Hvidberg et al., 2015; Jason et al., 2017).

To address the 21 diseases in the study, I pulled data from the Reddit API¹ searching by disease name and common synonyms².

This returned all submissions available in the database that had enough content in the title, post text, or comments, related to the disease name to be returned as relevant to that disease. Submissions that were returned both for a disease name and its synonym were deleted. Submissions that were returned for multiple disease conditions were deleted so that the model could be trained on unique submissions relevant to only one disease. Data for an additional five diseases was collected midway through the study to double check results. These additional diseases are those to the right of the red line in Figure 4.2 on page 49.

Once connected to the Reddit API, I wrote a script to collect all the posts related to each disease, all the posts' comments and all the sub-comments. A

¹Application Programming Interface, is a set of functions that allows a developer to access the data of an operating system or application

²Five diseases were dropped due to lack of data. These included Abdominal Wall Hernia, Chronic Renal Failure, Bleeding Ulcer, Heart Angina and Diabetes Type II. It is worth noting that some diseases are not represented in the data. Being that this is a forum where doctors are likely to complain, these diseases are thought not to present enough frustration for doctors to warrant posting here. They likely have relatively low stigma, with the exception of Diabetes Type II which did not have enough differentiating characteristics for the Reddit API to separate it from Diabetes Type I.

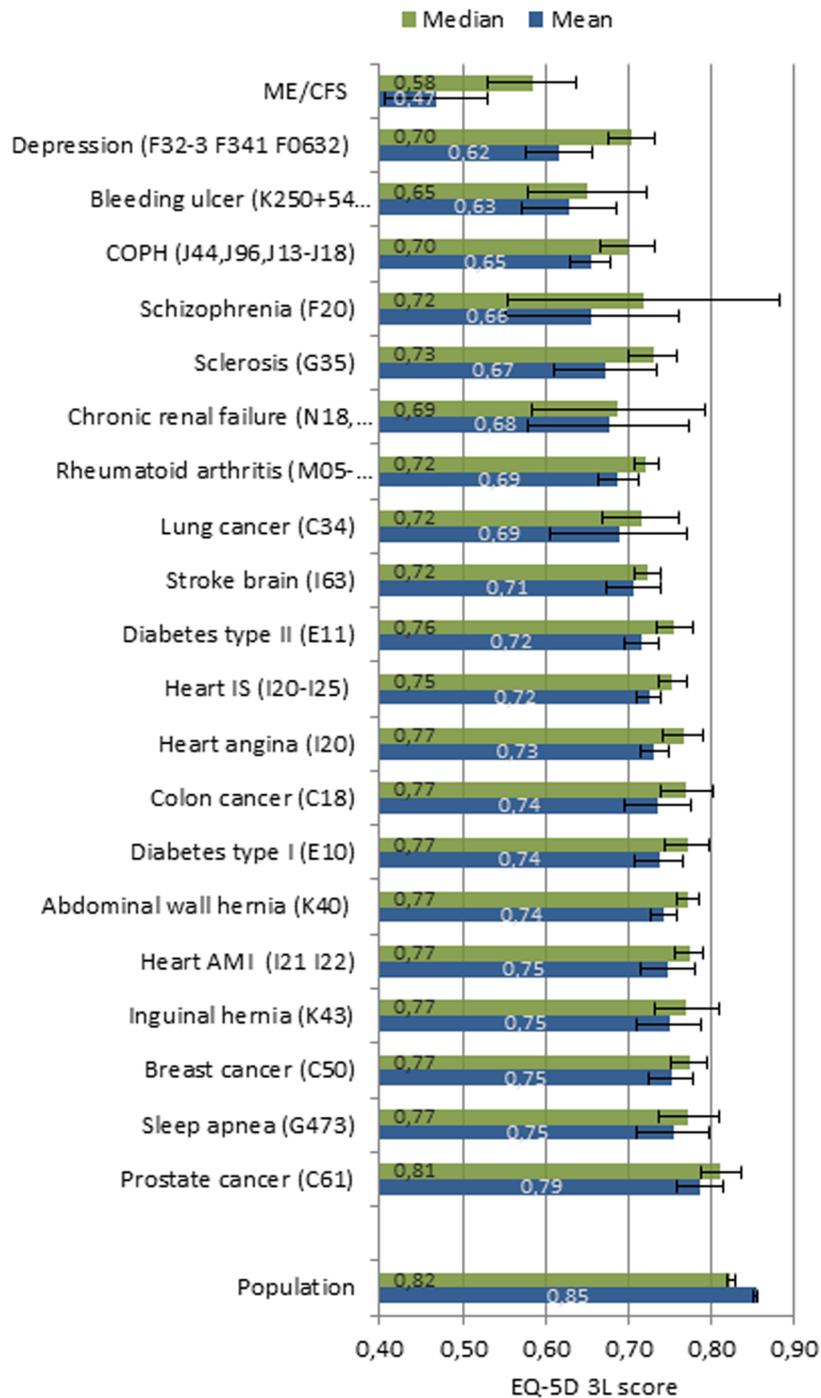


Figure 4.1: Mean and median Health-Related QOL compared for different conditions. ME/CFS patients on average had Health-Related QoL utility score of 0.47, or 0.56 after controlling for gender, age and co-morbidity using OLS. Controlling for other factors changed the level, but not the order of the conditions (Falk Hvidberg et al., 2015).

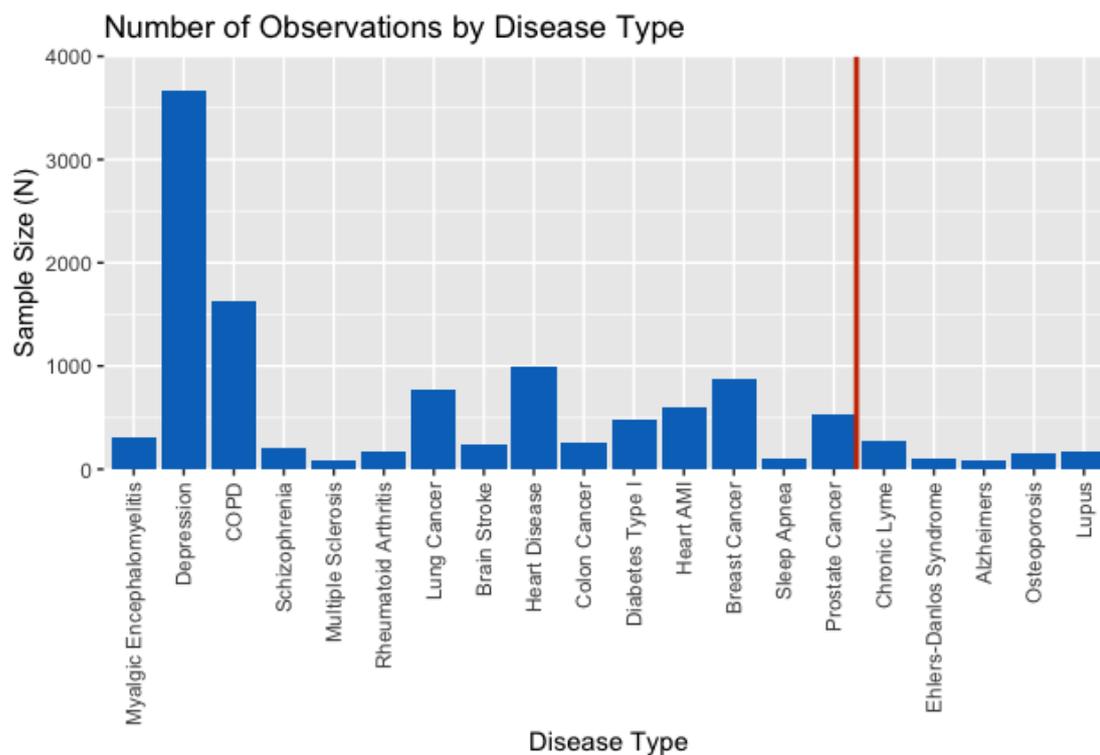


Figure 4.2: Number of documents (posts + comments) for each disease category. Diseases past the red line were added to test the validity of the Stigma dictionary discussed in Section 4.3 Methodology.

single post could have a couple of thousand comments branching off from a few hundred initial comments that broke into their own comment trees. The largest number of comments for any post in this data set was 800. Figure 4.2 shows the total number of observations for each condition.

For the posts and comments that were submitted, the following steps were taken to process the data:

- All text was converted to lowercase.
- URLs were found and removed using the regular expressions tools in Python. These allow the user to search for sequences of symbols within text data and remove them.
- Contractions were expanded.
- Common misspellings were found using a list of the most commonly misspelled words and were replaced with the corrections.
- All punctuation was removed.

- Stop words such as "a," "an," "of," and "the" were removed.
- Words that were repeats of disease names were removed. For example, so that "colon" would not be returned as a predictor for "colon cancer."
- Words were lemmatized, using a tool called WordNet³ lemmatizer that takes into account the context of the word and its morphology, or structure
- Text strings were split into tokens, or individual words, using the Natural Language Tool Kit, a Python library.
- A doc-term matrix was created from the tokens, which is a matrix that gives the frequency of words that appear in each post or comment. Words that appeared in less than five comments were deleted to remove any unusual misspellings not found by the spell check dictionary. The final matrix contained 11,128 observations of the 5,936 most common words. It follows an open vocabulary strategy as no pre-existing dictionaries were used (Schwartz, 2013).

These restrictions are necessary to help reduce the dimensionality of the data to make it more manageable (Gentzkow et al., 2017). The total sample size, N , is made up of 11,128 observations and includes 426 posts each with a title. Some of these posts are link-only posts and do not have a body containing any text. Meanwhile, 203 of them also have a body in addition to a title. The remaining 10,702 observations are comments. The body of each post and comment is considered its own document in the doc-term-matrix.

4.3 Methodology

In this paper, I design and implement a research strategy to test the hypothesis that natural language processing techniques can reveal doctors' attitudes toward specific medical conditions. This is the key contribution of this paper. I test this hypothesis by focusing on one disease, ME/CFS, and contrasting it with other diseases that share similar features. Investigating this required collecting novel data. I wrote a script to access the forum database and cleaned the returned data for

³<https://wordnet.princeton.edu/>

the purpose of this study. A literature review did not uncover any other papers in economics using data, text or otherwise, from this forum.

While unstructured data such as text can be analyzed qualitatively, I chose to use quantitative methods here. Due to the large number of words, the text was cleaned and processed using steps aimed at reducing dataset dimensionality. First, I divided the raw text, \mathcal{D} , into documents \mathcal{D}_i where each post and comment was one document and each document was labeled with the disease category used to search for it. The collection of all documents together, \mathcal{D} , is the corpus. The corpus \mathcal{D} was mapped to a numerical array \mathbf{C} , where \mathbf{c}_i is a row of \mathbf{C} . Each element of \mathbf{c}_i indicates the frequency of a word in document i . \mathbf{C} is also called a document-term matrix (Gentzkow et al., 2017).

Creating the document-term matrix, \mathbf{C} , required simplifying the grammar in the text by replacing words by their root. For example, jumping, jumps and jumped all have the same root form, jump. Replacing words with a common form reduces the feature space. The two ways of simplifying the form of words are stemming and lemmatization. Stemming replaces words by their root; hence, “economic,” “economics,” “economically” are replaced by the stem “economic.” It does not account for context, parts of speech or whether the stem itself is a word in the English language. Lemmatization takes more time and requires more complex algorithms, but accounts for the form of the word and more meaning is maintained. In the case of the word “saw,” stemming might return only “s” while lemmatizing would return “see.” For words such as “studies” and “study”, stemming would return “studi” and “study” while lemmatizing would return study for both (Manning, 2008; Bergmanis and Goldwater, 2018). Here, I lemmatized the words to take account of their forms and to improve interpretability.

Once the variables were transformed into a numerical document-term matrix \mathbf{C} , I standardized them and used LASSO-logistic regression to determine which words had the greatest predictive power for disease category. The algorithm for LASSO-logistic regression is shown in Equation 4.1. In traditional regression methods like OLS, the researcher usually has a clear idea about the model and the variables that

should be contained in it. In high-dimensional cases, with an X matrix that contain thousands of variables, it is not always clear which variables are relevant. LASSO can help by discarding variables that are not useful. In text data, there is usually a high degree of multicollinearity among the words while the LASSO penalty is indifferent to the choice among a subset of predictors that are correlated (Hastie et al., 2009).

The estimator for the LASSO-logistic regression is:

$$\hat{\beta}(\lambda) = \arg \min_{\beta} (n^{-1} \sum_{i=1}^n \rho_{\beta}(X_i, Y_i) + \lambda \|\beta\|_1) \quad (4.1)$$

and:

$$\|\beta\|_1 = \sum_{j=1}^p |\beta^j| \quad (4.2)$$

The loss function ρ from the equation above is:

$$\rho(\beta) = -y \left(\sum_{j=0}^p \beta_j x^j \right) + \log(1 + \exp(\sum_{j=0}^p \beta_j x^j)) \quad (4.3)$$

I used LASSO-logistic regression with binomial response variable to detect the words with greatest predictive power to determine whether a document is related to ME/CFS. This is a classification problem, which is a subset of machine learning problems with a discrete (or categorical) response variable. Here, the category is disease type. Further details regarding the benefits of LASSO and this approach to selecting features are in Appendix A.1. LASSO balances the bias-variance trade-off produced by the Expected Prediction Error shown in Equation 4.4.

$$\begin{aligned} \text{Expected Prediction Error} &= \mathbb{E} [(Y - \hat{f}(x))^2] \\ &= \text{Bias}^2 + \text{Variance}^2 + \sigma_{\epsilon}^2 \end{aligned} \quad (4.4)$$

where:

$\sigma_{\epsilon}^2 = \text{irreducible error}$

$Y = f(x) + \epsilon$ is a function with noise

as illustrated by (Hastie et al., 2009)

This application of LASSO returned a list of the most relevant words for ME/CFS on the forum as well as their coefficients. In an OLS regression the

coefficients show the rate of change in the response variable given a unit increase in the independent variable. With the logistic regression used here, the slopes of the coefficients show the change in log-odds⁴ of a comment being about ME/CFS for a unit change in the frequency of that word.

The logistic regression, which is detailed further in A.3, optimizes the sum of the likelihoods instead of the sum of squared residuals because logistic regression is solved using maximum likelihood. The binary model is used to determine whether a given comment is ME/CFS (1) or not (0). The multinomial version of this regression can be used to distinguish between multiple disease categories.

Next, I used the multinomial version of the LASSO-logistic regression. This regression has used a multinomial response variable to distinguish ME/CFS comments from comments of other diseases that share similar characteristics. A subset of the data was analyzed and the response variable was split into the categories ME/CFS, Depression and Autoimmune (MS/Lupus). In the previous case, there was a binary outcome variable indicating ME/CFS or not. In this case, the outcome variable had three categories.

The multinomial logistic regression model is derived by extending the binary case and is used when there are more than two unordered categories for the response variable. This application is correct for in this case because there is no way to determine which categories are more or less similar to ME/CFS. One example of ordered data would be responses that range from "strongly disagree" to "strongly agree." Depression was chosen because it is a mental illness that may share some stigma with ME/CFS due to being considered psychological. The purpose of including a mental illness was to see if ME/CFS discussions are likely to have more aggressive language above and beyond any stigma associated with mental illness patients. The two autoimmune diseases, Multiple Sclerosis (MS) and Lupus, were also included because of their severe effect on QOL and the symptom overlap between the illnesses (Jason et al., 2017).

⁴logarithm of the odds ($\frac{\pi}{1-\pi}$)

Using symptoms alone, Jason et al. (2011) found that it was difficult to distinguish ME/CFS from these two autoimmune diseases for epidemiological research. There is also a growing support for the theory that ME/CFS is an autoimmune disease, so it made sense to compare the disease as part of the same category (Nakatomi et al., 2014; Montoya et al., 2017; Davis, 2017). MS and Lupus are also accepted as physical illnesses which makes for helpful comparison with Depression.

For both the binomial and multinomial case, I used 5-fold cross-validation to select the model with optimal complexity. λ was chosen to maximize the model's predictive ability without overfitting the data. λ_{min} was chosen as the penalty parameter, which is the value for λ that gives the minimum mean cross-validated error. A description of the benefits of cross-validation as well as the algorithm used is available in Appendix A.2.

In order to further quantify this relationship, a new dictionary was created. Other sentiment analysis papers have used standardized dictionaries to measure tone, such as how positive or negative the text is using lists of common words. These dictionaries contain other common emotions such as Strength, Weakness, or Pain (Inquirer, 2002; Gentzkow et al., 2017). Simply measuring whether or not discussion is negative will not capture whether there is stigma because a negative statement about the patient's character will be counted as negative as would a negative statement lamenting the severity of a disease or the incidence of mortality.

One common dictionary, the General Inquirer, classifies over 8,000 words into 182 emotional categories. This dictionary was developed by Harvard Laboratory of Social Relations and is often used to analyze text for social science research (Inquirer, 2002). Within negativity, the General Inquirer divides sentiment further by labelling 833 of the words as Hostile to indicate an attitude of aggressiveness. The subset for Hostile contains words like "attack" or "accuse." This is not helpful for uncovering biases for the same reason that negativity is not. A doctor may express hostility toward their patients on the forum, but these cannot be separated from comments that use words like "attack" to refer to aggressively treating a disease. To

illustrate this, the Pearson correlation coefficients are shown between the General Inquirer dictionary and each of the diseases in the sample for four relevant emotions: Positive, Negative, Affiliation (or Supportive), and Hostile.

In order to summarize the level of stereotyping in the discourse, I manually classified words from the first 1,000 comments into a category of Stigma or not. Words that implied bias, disbelief, hostility, or frustration were labeled with a one for Stigma while all other words in the data were set to zero. In Wu (2017) the author classified words into the categories of personal or professional and regressed the gender of the posts on these two categories. Next, I checked the Pearson correlation coefficients between the classified stigma words and each of the diseases in the sample and tested whether those correlations were significantly different from zero. If the correlations were significant, then it would indicate a linear relationship between the Stigma dictionary and that disease.

Creating a dictionary for stigma provides several advantages. First, it would allow for the measurement of words that could show bias but were not chosen by LASSO. It is also a way of quantifying the differences in tone between diseases. Lastly, for the words that were picked up by LASSO for ME/CFS, the Stigma dictionary provides another way to be sure that these are not words used about other conditions in the sample; in particular, conditions that have similar characteristics.

Table 4.1: Frequency of Words in Each Category

	No. of Words	No. Added in New Sample
Custom Stigma Words	1244	50
General Inquirer	185,800	5,366
General Inquirer Hostile	14,643	417
Total Words Overall	263,399	7,884

The frequencies above are the number of occurrences of the words in each dictionary within the data set.

Different to classifying words into categories of Personal or Professional, choosing whether a word should be counted as Stigma may be considered to be a more subjective selection process. In order to minimize subjective input, the Stigma

dictionary was cross-referenced with a common dictionary used to analyze tone, while any words not in this dictionary were deemed excessively context specific and removed. For consistency, the General Inquirer dictionary was again used for this selection. This dictionary contains almost all English words that appear four or more times per million but does not contain proper nouns or domain-specific words. As the General Inquirer words were chosen by an independent team, use of this dictionary should reduce subjectivity. This goal was also furthered by the exclusion of rare words.

To be objective, there should be separation between the choice and classification of words and the analysis. By viewing the data while conducting the analysis, this approach could influence the word selection and classification process. A separation could be introduced by having separate researchers create the dictionary and classify the emotional tone from those who analyze the data. Alternatively, people without a natural language processing background could be paid to classify the emotions in text.

To address this source of influence, I collected data for an additional five diseases containing a mix of stigmatized and established diseases, then checked the correlations with the two Stigma dictionaries ⁵. The new diseases were: Chronic Lyme, Ehlers-Danlos Syndrome, Alzheimers, Osteroporosis, Lupus. Chronic Lyme and Ehlers-Danlos Syndrome were chosen because they appeared in the data as diseases with a high degree of hostility directed toward them, but they were not in the original sample.⁶ I chose Alzheimers and Osteroporosis because they are diseases that affect older patients who may be victims of ageism. The intention was to check whether or not the hypothesized conditions (ME/CFS, Chronic Lyme and EDS) had stigma in their tone that could be detected following a different pattern from potential ageism. In this way, a blinded test of the correlation could be run checking for Stigma in the tone without viewing any of the text data for those conditions.

⁵The custom Stigma dictionary and the smaller Stigma dictionary with only General Inquirer words

⁶Preliminary research shows that Chronic Lyme is currently diagnosed in more women than men (Wormser and Shapiro, 2009).

4.4 Word Selection

The model in this paper aims to predict the disease classification of a post or comment using the frequency of the most common words, excluding the names of the diseases themselves. The model aims to determine which words in forum posts have the strongest predictive power for each disease-type. The frequencies of the words are indicated by x_i . The posterior probabilities in this case are of a comment being categorized as ME/CFS (π) or not ($1 - \pi$). These probabilities are:

$$\pi_i = P(ME_i = 1|x_i) = \frac{\exp\{\beta_0 + x_i^T \beta\}}{1 + \exp\{\beta_0 + x_i^T \beta\}} \quad (4.5)$$

$$1 - \pi_i = P(ME_i = 0|x_i) = \frac{1}{1 + \exp\{\beta_0 + x_i^T \beta\}} \quad (4.6)$$

The likelihood of any comment being an ME/CFS comment (shortened to ME for convenience):

$$P(ME_i|x_i) = P(ME_i = 1|x_i)^{ME_i} \times P(ME_i = 0|x_i)^{1-ME_i} \quad (4.7)$$

Assuming independence of observations, the log-likelihood of n observations becomes:

$$l_n(\beta) = \log(\prod_{i=1}^n P(ME_i|x_i)) = \sum_{i=1}^n ME_i(\beta_0 + x_i^T \beta) - \log(1 + \exp(\beta_0 + x_i^T \beta)) \quad (4.8)$$

The objective function then used to estimate β is:

$$\hat{\beta}(\lambda) = \arg \min_{\beta} \{(-l_n(\beta)) + \lambda \|\beta\|_1\} \quad (4.9)$$

and:

$$\|\beta\|_1 = \sum_{j=1}^p |\beta^j| \quad (4.10)$$

The LASSO-logistic model determines which words have the most predictive power for disease-category. This method takes account of word frequency so that more common words within a disease-category, such as disease-specific treatment types, have higher probabilities, but words that appear frequently across all diseases,

such as medicine or patient, are weighted down. Each x_i is a k by 1 vector of word counts that can be stacked into the X -matrix which is $n \times k$ doc-term-matrix.

Words that are not selected by the model have coefficients of zero. These words have been determined to be irrelevant to the disease-type of the post. Where x_{ik} is the number of occurrences of word k in comment i and $x_{i,(-k)}$ is the vector of words in comment i not including k . For the remaining words, the marginal effect of the k th word is the difference between the probability of having one extra occurrence of that word and having the actual number of occurrences:

$$\begin{aligned} & P(ME_i = 1 | x_{i,(-k)}, x_{i,k} + 1) - P(ME_i | x_{i,(-k)}, x_{i,k}) \\ &= \frac{1}{n} \sum_{i=1}^n \{P(ME_i = 1 | x_i) \times P(ME_i = 0 | x_i)\} \hat{\beta}_k \end{aligned} \quad (4.11)$$

The coefficients express the marginal effect of each word selected to be included in the model, representing the change in log-odds that the comment is related to ME if the frequency of the word increases by 1 (Leeper, 2018).

The same way the patients get ostracized, the physicians get ostracized. Before I got into CFS, I had a very comfortable career, but in order to be seen with the same validity I had with other illnesses, I had to double my effort. If I send a paper on Toxoplasmosis, 95% of journals will take a look. If I send a paper on CFS, although I have the same name, 100% of the journals will reject the first time. 100%. Then I have to go through the whole process again.

Dr. Jose Montoya, Infectious Diseases Professor, Stanford University

5

Results

5.1 LASSO-Logistic Regression - Binomial Case

In Table 5.1, we can see the words with the greatest predictive power for whether or not a given post is related to ME/CFS. Words that are part of the Stigma dictionary are in bold and words that are medical terminology are in italics. An attempt was made not to classify commonly understood terms as medical and to only include technical medical terms and jargon. The stigma dictionary was made using the top 1,000 comments, tagging all the words that have to do with stigma or indicate disbelief. These words are only indicated in bold for now, but this dictionary is used explicitly in the next section. The medical terminology words in italics are indicated only for clarity.

In an OLS regression, the coefficients would show the rate of change in the response variable given a unit increase in the independent variable. Here, the slope is the marginal change in the log-odds¹ of the response variable. This is approximately equal to the ratio of probabilities. If $\exp^{\beta_k} = 2$ then a unit change in x_k makes a comment twice as likely to be about ME/CFS. $\exp^{\beta_k} = 2$ could be considered to be a 100% increase in relative risk as $\exp^{\beta_k} - 1 = 1$. Since the

¹logarithm of the odds ($\frac{\pi}{1-\pi}$)

Table 5.1: Words with Strongest Predictive Power for ME/CFS - Binomial

name	coefficient	name	coefficient
afflict	3.64	madeup	0.88
skeptic	3.64	psychological	0.79
jamison ^a	3.62	saga	0.78
unproductive	2.98	<i>epidemiologists</i>	0.76
<i>sibo</i>	2.83	plurality	0.76
<i>sds</i>	2.63	<i>cidp</i> ^a	0.76
nonpsychiatric	2.16	scoff	0.73
morgellons	2.10	researchers	0.69
towel	1.97	somatized	0.53
<i>candida</i>	1.85	agenda	0.38
scientists	1.71	<i>cis</i>	0.30
reproducible	1.69	<i>ferritin</i>	0.27
intolerance	1.59	conspiracy	0.26
overgrowth	1.51	relay	0.22
<i>pulmonologist</i>	1.50	norway	0.20
instagram	1.45	illnesses	0.19
psychosomatic	1.33	vague	0.17
documentary	1.33	<i>aerobic</i>	0.14
disprove	1.30	<i>nerve</i>	0.10
<i>mudphud</i>	1.30	stanford	0.08
exertion	1.26	post	0.07
elude	1.11	science	0.06
nothingness	1.07	disable	0.02
<i>hypermobile</i>	0.88	savella	0.01

^aJamison Hill is an author and famous ME/CFS activist who often writes for The New York Times

^aChronic inflammatory demyelinating polyneuropathy - a neurological condition comorbid in many ME/CFS patients

Table lists the features chosen by LASSO-logistic model as having the greatest predictive power for whether a given post is about ME/CFS and lists the coefficient size. The 5-fold cross-validation used to choose value for penalty parameter λ . Regression was binomial with a single outcome variable of 0 or 1 indicating if the post was about ME/CFS. Words in **bold** are part of the Stigma dictionary. Words in *italics* are medical terminology.

variables have been standardized, the coefficients show the change in log-odds for an increase of one standard deviation in frequency for a given word.

For ME/CFS, three words with the strongest predictive power are "skeptic," "unproductive," and "mogellons."² The coefficients are 3.64, 2.98, and 2.10 and finding exp^{β_k} for each coefficient gives 38.09, 19.68, and 8.16. A one standard deviation increase in the frequency of the word "skeptic" will multiply the relative risk of a comment being about ME/CFS by 3809%. Phrased another way, this is a 3709% increase in relative risk of a comment being about ME/CFS compared to any other condition. For "unproductive" and "morgellions" the increases in relative risk are 1868% and 716%

5.2 LASSO-Logistic Regression - Multinomial Case

These results were repeated with the multinomial logistic regression. Due to the large number of diseases, it was not possible to perform cross validation on a multinomial regression with all the diseases. Instead, a subset of the diseases was analyzed: ME/CFS, Depression and Autoimmune Disease. The reason for including depression was that it would allow for comparison between ME/CFS and another potentially stigmatized condition. It would also reveal possible differences in language between ME/CFS and mental illness. The two autoimmune diseases, MS and Lupus, were added into their own category for comparison because ME/CFS has a similar set of symptoms to both MS and Lupus; additionally, the three diseases are of similar severity (Jason et al., 2011, 2017).

The multinomial model generalizes the binary case. The difference between the multinomial regression and the binomial regression is that, in the binomial, there are only two outcomes for the response variable: ME/CFS comment or not. For multinomial regression, there are three outcomes: ME/CFS, Depression, or Autoimmune. One key assumption for the multinomial regression is that the outcomes are nominal, there is no order to the findings and it cannot be said that

²MedicineNet describes Morgellons disease as "a delusional disorder that leads to the belief that one has parasites or foreign material moving in, or coming out of, the skin"(Roxanne Dryden-Edwards, MD, n.d.).

one is greater than the other. Outcomes are ordinal if they can be arranged in a meaningful way as in from “excellent” to “very poor”. If the outcome is something like cities, choice of major or disease type, then there is no apparent ordering.

For a multinomial regression with M possible outcomes, there need to be $M - 1$ independent binary logistic regression models, where one class is chosen as the base and the other $M - 1$ classes are regressed against the base class. As there are three classes for this regression, two binary logistic regression models are required, which can be derived from a binomial regression where one class is chosen as the reference class and others are set to zero. This step is then repeated for $M - 1$ categories.

The probabilities for the multinomial regression are slightly different. For the reference category:

$$P(Y_i = 1) = \frac{1}{1 + \sum_{m=2}^M \exp(Z_{mi})} \quad (5.1)$$

where Z_i ensures the probabilities form a proper probability distribution and sum to one.

$$Z_i = \sum_{m=1}^M P(Y_i = m) = 1 \quad (5.2)$$

implies that:

$$Z_i = \sum_{m=1}^M \exp(\beta_m X_i) \quad (5.3)$$

The probability of a word being in each of the other categories is compared to the reference category. For a response variable with M categories, there are $M - 1$ equations relative to the reference category. For each of the $M - 1$ cases, there will be a predicted log-odds equation. The relationship between each category and the reference category is:

$$Z_{mi} = \ln \frac{P(Y_i = m)}{P(Y_i = 1)} = \beta_{m0} + \sum_{k=1}^K \beta_{mk} X_{ik}, = \beta_{m0} + \beta_m^T X_i = m = 2 \dots M \quad (5.4)$$

where:

$k = \text{number of words}$

$i = \text{number of documents} = \text{posts} + \text{comments}$

β_{mk} is the coefficient for the k th explanatory variable and the m th category.

This simplifies to:

$$P(Y_i = m) = \frac{\exp(Z_{mi})}{1 + \sum_{m=2}^M \exp(Z_{mi})} \quad (5.5)$$

The multinomial logistic model has M coefficient vectors of log-odds parameters $\beta_1, \beta_2, \dots, \beta_M$ and β_m is a vector with k values. The irrelevant parameters are forced to 0, while maximizing the penalized log-likelihood:

$$\max_{\{\beta_{0m}, \beta_m\}_1^M} \left\{ \sum_{i=1}^n \log P(ME_i | x_i) - \lambda \sum_{m=1}^M \sum_{k=1}^K |\beta_{mk}| \right\} \quad (5.6)$$

The multinomial model makes the same assumptions as the binary case except has $M - 1$ categories. These assumptions are:

- Dependent variable is measured at the nominal level
- Dependent variable has mutually exclusive and exhaustive categories
- Independence of irrelevant alternatives

Independence of irrelevant alternatives requires that the ratio of shares is constant between dependent variables. So if one category is removed, then every other category experiences a proportional change that makes the relative shares the same. If the original split was $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$, then when one category is removed the probabilities will become $\frac{1}{2}, \frac{1}{2}$.

The results for the ME/CFS words in the multinomial case, shown in Table 5.2, were similar to the results in the binomial case. Several words from the Stigma dictionary were chosen as systematically predictive of ME/CFS including “laughable,” “denigrate,” “snarky,” and “made-up.” There is medical terminology in the predictive words list for ME/CFS, but the language is not extensively technical. The majority of the medical words could be understood by someone without a medical degree such as “pulmonologist” and “mutations.” ME/CFS has 17 predictive words in italics indicating medical jargon, compared to only 6 for depression. ME/CFS, compared to depression, has nearly three times more medical words with predictive power. ME/CFS has more words selected to have predictive power than depression does, but in both cases, this is out of the total number of words chosen by the LASSO function. Despite depression having fewer total words

Table 5.2: Words with Strongest Predictive Power for **ME/CFS** in Subset of Data: ME/CFS, Depression and Autoimmune - Multinomial

name	coeff	name	coeff	name	coeff
worldview	3.80	intolerance	1.59	flex	0.85
<i>mitochondria</i>	3.66	combat	1.58	surely	0.84
redundant	3.36	exertion	1.57	science	0.83
abnormality	3.32	degrade	1.55	overstate	0.83
donate	3.14	mutation	1.54	madeup	0.77
relay	3.06	wider	1.54	detrimental	0.73
<i>microbiome</i>	3.04	snarky	1.39	dismiss	0.72
<i>ulcers</i>	3.00	illustrate	1.36	scientifically	0.72
intestine	3.00	<i>cis</i>	1.36	norway	0.66
midline	2.91	<i>hypermobile</i>	1.35	user	0.66
randomize	2.83	simplistic	1.35	dishearten	0.64
instagram	2.72	akin	1.32	disable	0.63
misdiagnosis	2.64	shill	1.30	<i>pcps</i>	0.61
crystal	2.56	reverse	1.29	research	0.57
<i>pulmonologist</i>	2.40	rename	1.27	summarize	0.56
morgellons	2.38	psychosomatic	1.25	subacute	0.54
disprove	2.26	<i>picc</i>	1.23	sponsor	0.54
laughable	2.24	accident	1.20	psychological	0.54
denigrate	2.21	reproducible	1.20	character	0.53
<i>immunotherapy</i>	2.11	harmful	1.19	unremarkable	0.51
anaerobic	2.07	toss	1.13	pace	0.50
nonpsychiatric	2.06	olive	1.11	logic	0.49
<i>emg</i>	2.05	unprofessional	1.11	valid	0.47
null	2.03	<i>huntingtons</i>	1.10	proportion	0.46
probiotics	1.97	<i>septic</i>	1.07	illnesses	0.45
observations	1.96	<i>mutations</i>	1.07	researchers	0.44
centuries	1.87	traumatic	1.03	planet	0.43
<i>pathologic</i>	1.81	garden	1.02	heavy	0.43
<i>ferritin</i>	1.80	mediocre	0.99	agenda	0.41
citation	1.74	disappoint	0.98	criticism	0.40
lens	1.74	consume	0.98	:	:
independently	1.73	threat	0.98	skeptical	0.39
scientists	1.72	unaware	0.91	refuse	0.37
teenage	1.70	surveillance	0.91	invisible	0.26
<i>neurological</i>	1.69	<i>biomedical</i>	0.91	legitimate	0.24
documentary	1.63	turf	0.87	bias	0.20
workups	1.61	film	0.87	secondary	0.06
intrinsic	1.60	officer	0.86		

Table list the features chosen by LASSO-logistic model as having the greatest predictive power for whether a given post is about ME/CFS and lists the coefficient size. The 5-fold cross-validation used to choose value for penalty parameter λ . Regression was multinomial on a subset of the data with just these three disease categories. It had a categorical outcome variable indicating whether the post was about ME/CFS, depression, or autoimmune disease. Words in **bold** are part of the Stigma dictionary. Words in *italics* are medical terminology.

Table 5.3: Words with Strongest Predictive Power for **Depression** in Subset of Data: ME/CFS, Depression and Autoimmune - Multinomial

	name	coefficient		name	coefficient
	months	0.33		<i>suicide</i>	0.09
	<i>ketamine</i>	0.31		risk	0.08
	<i>surgery</i>	0.28		program	0.08
	dose	0.28		things	0.08
psychiatrist	psychiatrist	0.28		impossible	0.08
	work	0.27		top	0.08
	issue	0.27		management	0.07
	fail	0.25		exactly	0.07
	<i>medication</i>	0.25		<i>drug</i>	0.07
	stress	0.22		definition	0.06
	time	0.22		drink	0.06
	<i>psych</i>	0.20		echo	0.06
	tell	0.18		meet	0.05
	great	0.17		weight	0.05
	recommend	0.17		year	0.05
	effect	0.17		options	0.05
	insurance	0.16		monitor	0.04
	place	0.16		<i>ill</i>	0.04
	state	0.15		odd	0.04
	lose	0.15		death	0.04
	consider	0.14		end	0.03
	die	0.14		hours	0.03
	leave	0.14		option	0.03
	mine	0.13		stable	0.02
	care	0.11		helpful	0.02
	kill	0.11		massive	0.02
	question	0.11		worth	0.02
	sound	0.11		easier	0.02
	wonder	0.11		rat	0.02
	job	0.11		pay	0.02
	feel	0.11		practice	0.02
	help	0.11		license	0.01
	live	0.10		honestly	0.01
	completely	0.10		ventricular	0.01
	require	0.09		reason	0.01

The table list the attributes with the greatest predictive power for whether a given post is about ME/CFS and lists the coefficient size. A 5-fold cross-validation is then used to choose value for penalty parameter λ . I used multinomial regression on a subset of the data with just these three disease categories. It had a categorical response variable indicating whether the post was about ME/CFS, depression or autoimmune disease. Words in **bold** are part of the Stigma dictionary. The explicit medical terminology is in *italics*.

Table 5.4: Words with Strongest Predictive Power for **Autoimmune Diseases (MS and Lupus)** - Multinomial

name	coeff	name	coeff	name	coeff
unchanged	3.84	entertainment	1.89	<i>vasculitis</i> ^a	0.81
metal	3.43	band	1.82	disruption	0.78
<i>ibd</i> ^a	3.36	<i>plaque</i>	1.80	inspiration	0.76
markup	3.28	<i>spine</i>	1.79	<i>ppms</i>	0.76
intelligence	3.26	classification	1.74	<i>os</i>	0.73
<i>ol</i> ^b	3.03	<i>icd</i>	1.71	newly	0.71
weed	3.01	coordination	1.69	succeed	0.71
conduction	3.00	bloody	1.64	premature	0.67
<i>bnp</i> ^c	3.00	<i>parkinsons</i>	1.56	<i>dna</i>	0.67
fevers	3.00	<i>leukemia</i>	1.54	<i>femoral</i>	0.66
<i>effusions</i>	3.00	viruses	1.53	<i>connective</i>	0.63
venous	3.00	<i>intermittent</i>	1.52	killer	0.55
<i>lipid</i>	2.95	topics	1.47	<i>platelet</i>	0.43
<i>endocarditis</i>	2.83	game	1.44	asses	0.41
pus	2.74	hashimotos	1.42	price	0.39
<i>cardioversion</i>	2.69	<i>virus</i>	1.38	fare	0.37
<i>lymphoma</i>	2.55	<i>fibrosis</i>	1.34	confidently	0.34
<i>esophageal</i>	2.55	<i>rectal</i>	1.28	<i>asthma</i>	0.34
<i>neuropathic</i>	2.33	ring	1.15	<i>melanoma</i>	0.32
convenient	2.31	growth	1.12	midlevels	0.30
spell	2.28	<i>vivo</i> ^a	1.02	overboard	0.29
<i>biliary</i>	2.21	<i>albumin</i> ^b	1.00	plug	0.17
cartilage	2.17	<i>smear</i>	0.99	<i>antivirals</i>	0.16
<i>nsclc</i> ^d	2.15	paranoid	0.98	rural	0.16
<i>pap</i>	2.10	<i>gestational</i>	0.97	<i>immune</i>	0.13
<i>pimp</i>	1.98	<i>methotrexate</i> ^c	0.96	<i>cystic</i>	0.11
incompetence	1.97	<i>celiac</i>	0.95	demonstrate	0.07
<i>vaccine</i>	1.96	humor	0.92	rice	0.07
<i>malignancy</i>	1.93	doable	0.90	<i>rituximab</i>	0.05
odyssey	1.93	modify	0.89	classic	0.05
<i>ac</i> ^e	1.90	<i>precordial</i> ^d	0.83	storage	0.05
				psychology	0.02

^aIritable Bowl Disease^bleft eye, used in writing prescriptions^cBrain natriuretic peptide^dnon-small-cell lung carcinoma

noma

^ebefore meals^amedical device company^bprotein made by liver^cimmune system drug^dportion of body over^ainflammation of blood vessels

The table lists the attributes with the greatest predictive power according to whether a given post is about ME/CFS and lists the coefficient size. Then, 5-fold cross-validation was used to choose value for penalty parameter λ . Multinomial regression was used on a subset of the data with these three disease categories alone. This regression had a categorical response variable determining whether or not the post was about ME/CFS, depression, or autoimmune disease. Words in **bold** are part of the Stigma dictionary. Explicit medical terminology is in *italics*.

in the list, the difference is still significant; LASSO has chosen all of the words that systematically differentiate between the diseases.

Comparing the results for ME/CFS in Table 5.2 to those for depression in Table 5.3, depression has fewer stigma words. There is only one word from the Stigma dictionary, "psychiatrist," and the coefficient is .28. This coefficient is lower than the most of the coefficients for ME/CFS. There were no stigma words in the list of selected predictive words with coefficients below .02. Compared to ME/CFS which has 22 stigmatizing words, depression has only one. Depression also has very few words indicating medical terminology. Arguably all of the words with predictive power for depression could be easily understood by someone with no medical background.

The results for ME/CFS in Table 5.2 are also much different to those for autoimmune diseases in Table 5.4. For the autoimmune category, there are fewer words from the Stigma dictionary chosen as predictive than for ME/CFS. The autoimmune category shows two words from the Stigma dictionary as having predictive power compared to 22 for ME/CFS. There are 11x more stigma words chosen as highly predictive for the ME/CFS category than for autoimmune disease. Words that have medical meaning, especially particularly technical terms, are again in italics. There is much more medical terminology in the autoimmune list including drug names, jargon used for writing prescriptions, and other words that have to do with body systems and disease processes. There are 47 words indicated as medical jargon for this category; this is 2.7x more words than for ME/CFS. The words are also noticeably more technical than for the other two categories; there are fewer words that a person who is not medically trained would understand. The more numerous presence of medical terminology indicates greater engagement with the disease process and less discussion regarding presumed patient personality.

5.3 Dictionary Correlations

Next, I checked the correlations between four relevant emotions in the General Inquirer dictionary and the frequencies of words for each condition. Then, I did the same analysis for the custom Stigma dictionary that I had created and a second Stigma dictionary created from the overlap of the custom Stigma words and the General Inquirer dictionary.

In Table 4.1 on page 55, the dictionaries used for the correlations analysis are shown alongside the frequencies of words from each dictionary category. The frequencies are separated into two categories: the number of words in the initial sample of diseases and the number of words in the second sample containing the five additional diseases. Table 5.5 lists examples of words from each dictionary used. Stigma words that did not appear in the General Inquirer dictionary proved to be less common words, which may be very specific to this context. The purpose of using only stigma words that do appear in the General Inquirer dictionary is to have a methodology for testing for stigma that reduces potential bias.

Table 5.6 on page 74 shows the correlations between the diseases in the sample and four relevant emotions from the General Inquirer dictionary: Positive, Negative, Affiliation (Supportive), and Hostile. These are correlations between a vector of the frequency counts of every word for that disease, and a vector of zeros and ones indicating whether or not the word was tagged with a given emotion. The correlations were only between the words that were in both lists: the frequency matrix from the data and the 8,719 words in the General Inquirer dictionary.

The Pearson correlation coefficients shown in Table 5.6 are:

$$\rho_{\hat{\alpha}, \hat{\beta}} = \frac{\text{cov}(\hat{\alpha}, \hat{\beta})}{\sigma_{\hat{\alpha}} \sigma_{\hat{\beta}}} \quad (5.7)$$

The square of this Pearson correlation coefficient is equal to the R^2 that would come from the OLS regression between the two variables, term frequency and sentiment where an intercept is included. For the first row of Table 5.6, 0.05 is the correlation coefficient of the word frequencies for ME/CFS with a binary variable indicating 1 if the word is in the General Inquirer dictionary for Positive and 0 if it is not. The

Table 5.5: Dictionary Used for Sentiment Analysis - Total Words Contained and Examples from Each

Dictionary		Total Words	Examples
General Inquirer		8719	common, curious, diversion, dress, dreamer, eliminate, hungary, intervene, italy, jazz, keep, literal, rest, team, taught, undoubtably, unworthy, up, ubeat, whimper, whole, yesterday, zeal
General Inquirer Hostile		833	abhor, abolish, abrasive, abuse, attack, belittle, betray, disbelief, disgust, lying, malicious, malice, stubborn, unwillingness, violent
Stigma	Stigma also in General Inquirer (maintained more objective words)	62	assumption, attention, bizarre, bogus, bunk, conspiracy, crazy, deception, degrade, dismissive, dramatic, faking, frustration, hysterical, imagine, joke, lazy, liar, magical, misinformation, mysterious, mysteriously, presumption, ridiculous, scoff, skeptical, subjective, suspicious, trick, unimportant, unsound, useless, vague
	Stigma not in General Inquirer (removed less objective words)	21	agenda, crazies, denigrating, entitled, laughable, made-up, magically, malingering, misconception, misinterpret, placebo, psychosomatic, self-diagnose, skeptic, somatized, undiagnosable, unexplained, wacko

General Inquirer row lists examples from the General Inquirer dictionary. The General Inquirer Hostile row lists examples of the subset of General Inquirer words that are tagged with the emotion category of hostile. The custom stigma dictionary is broken up into two parts: 1) examples from the stigma dictionary created for this study and 2) examples of the stigma words that do not appear in the General Inquirer dictionary. Stigma words that do not appear in General Inquirer were removed to reduce subjectivity.

p-value is smaller than 0.00, indicating that it is statistically significant past the 99% significance level. ME/CFS is also correlated with Negative and Hostile tone but not with Affiliation (Supportive). Similar to ME/CFS, many of the diseases in the dataset were correlated with emotions from the General Inquirer dictionary. Those that have *t* – statistics above 1.96 are indicated in bold.

It is possible that diseases with negative stereotypes have more negative words. If this were true, it would be indicated in the correlation with the dictionary for negativity. The problem with this approach is that "these patients are a pain," and

"these patients are in pain," are both negative and would both be tagged with the same emotional intensity. One way around this is to use the sub-dictionary for hostility, but this runs into similar issues. Words like "deceit" are in the dictionary of hostile words, but so are words like "attack" and you can also attack a disease.

To work around this, a new dictionary was created to test for stigma directly. The words from the top 1,000 comments were tagged as having stigma or not. As expected, this new dictionary was correlated with ME/CFS and not with any of the other diseases. However, as the data had already been viewed, there could be bias introduced during the dictionary creation.

This could be resolved by using different researchers for the stigma dictionary creation from those running the analysis. Instead, data was collected for five more diseases in order to complete a blind test. The new diseases were: Chronic Lyme, Ehlers-Danlos Syndrome (EDS), Alzheimer's, Osteoporosis and Lupus. These include two diseases that I hypothesized to be stigmatized (Chronic Lyme and EDS) and three diseases I hypothesized would not be (Alzheimer's, Osteoporosis and Lupus). Then, a blinded test of the correlations was run on those diseases where none of the data had been viewed previously.

In a given post or comment, there are likely to be very few stigma words. For example, if a patient is called "lazy" then that word may show up only once in the comment, while the presence of this word could be dampened by the presence of many other non-stigma words in the comment. For this reason, if the Stigma dictionary was not correlated with the word frequencies, it would be useful to compare differences between diseases. Yet this outcome did not turn out to be the case. The new Stigma dictionary was significantly correlated with whether a document was ME/CFS and not with the other diseases in the sample.

Out of the additional diseases, the custom dictionary was correlated with one of the two diseases that were expected to be stigmatized (Chronic Lyme, but not EDS) and none of the 3 diseases expected not to be stigmatized (Alzheimer's, Osteoporosis, and Lupus). The results of this check for correlation with the Stigma dictionary are in the first column of Table 5.7. The coefficients are lower than those

from the General Inquirer dictionary in Table 5.6 because fewer words were tagged with stigma than with any of the emotions in the General Inquirer dictionary.

The coefficients are also unlikely to be significant because it is a simple test for correlation between the frequency of the words and whether or not they have a given emotion. For stigma, even if a text has many problematic words, these words would be used at most once in a sentence or once in an entire post. They will be drowned out from the noise of all the other words in the post.

One way of getting around this problem would be to look only at words that are selected by LASSO to narrow down the total words examined, or to look at differences in means between groups. However, this approach was not necessary because the correlation was revealed significant at this level, although it remained small in size because there are so many other words. There is an association here although it is a small part of the total words. This information is hence still of some value because the Stigma dictionary was significantly correlated with the diseases predicted to be subject to incorrect biases, but not with the diseases predicted not to face this bias.

To reduce subjectivity further, the custom Stigma dictionary was joined with the General Inquirer dictionary. Any words that did not appear in the General Inquirer dictionary were removed, which would eliminate words that are particularly context specific. When the correlations were run with the generic General Inquirer Stigma words, the result stands. These results are in the second column of Table 5.7.

The original stigma dictionary is correlated with both ME/CFS and Lyme and not with the other diseases. Using the second stigma dictionary that included only the overlap of the Stigma and General Inquirer words, stigma words were still significantly correlated with whether a document was ME/CFS, and not with the other diseases in the sample. The significance level for the correlation with Lyme goes down from 5% to the 10% significance level using the more generic stigma words. The results for EDS were contrary to expectations in that the frequencies were not significantly correlated with either of the two Stigma dictionaries.

For brevity, Table 5.7 reports only the Pearson coefficient (square root of R^2) and the statistical significance. These are the same values that would result from a univariate regression of word frequency on stigma words. The outcome variable, word frequency, is a list of every word for each disease and the number of times it occurs across all the documents in the corpus. The other diseases are filtered out, entailing that only one disease is considered at a time. Each word is an observation while the independent variable, stigma words, is a binary variable indicating 1 if the word is in the Stigma dictionary and 0 if it is not.

To see in more detail the diseases that had significant t – *statistics*, the full OLS regressions are reported in Table 5.8. In Table 5.8, the coefficients and intercepts from the regressions are reported in addition to the R^2 values and statistical significance. For purposes of comparison, between the diseases that were significant and not, I added one disease to Table 5.8 which was not significantly related to the stigma words, Ehlers-Danlos Syndrome.

With a binary variable as the regressor, these standard errors will not be unbiased but they remain asymptotically unbiased. The number of words is technically discrete, though approximately continuous. In large samples such as the one here, the standard asymptotic results apply. We can hence treat the standard errors as belonging to an approximately normally distributed variable.

Table 5.8 shows that stigma words appear on average more frequently than the average for all other words. For ME/CFS, if a word is in the custom Stigma dictionary, then it is likely to have a frequency of 2.4 additional words than the average non-stigma word. The constant is .8 here indicating that non-stigma words show up on average .8 times for ME/CFS. The combined Stigma/General Inquirer dictionary was associated with 1.6 additional words than words that were not in the dictionary. This is lower because context-specific words have been filtered out and only generic words remain. However, the value still has a significant t – *statistic*. For Chronic Lyme, the word frequencies were lower - .9 and .6 for the Stigma and Stigma/General Inquirer dictionary - but remained significant. For EDS, these effects were smaller and did not have statistical significance just as before.

One may anticipate that medical professionals are objective and treat all conditions with urgency. However, this analysis has shown that doctors' attitudes are inconsistent across disease type. Attitudes differ in relation to separate diseases. This finding holds when comparing diseases with similar severity and symptom profiles. Through a combination of LASSO-logistic regression and checks for correlation, the results have shown that, when measured by word choice, the attitudes expressed by medical providers as enable conditions to be distinguished.

Table 5.6: Correlations between Word Frequency by Disease Category and General Inquirer Dictionary

Disease	Positive	Negative	Affiliation	Hostile
Myalgic Encephalomyelitis	0.05 (0.00)	0.03 (0.00)	0.06 (0.00)	0.02 (0.03)
Depression	0.09 (0.00)	0.04 (0.00)	0.09 (0.00)	0.04 (0.00)
COPD	0.09 (0.00)	0.02 (0.08)	0.07 (0.00)	0.03 (0.00)
Schizophrenia	0.06 (0.00)	0.02 (0.03)	0.05 (0.00)	0.01 (0.29)
Multiple Sclerosis	0.05 (0.00)	0.02 (0.06)	0.04 (0.00)	0.02 (0.02)
Rheumatoid Arthritis	0.05 (0.00)	0.03 (0.00)	0.04 (0.00)	0.01 (0.16)
Lung Cancer	0.09 (0.00)	0.02 (0.03)	0.10 (0.00)	0.03 (0.01)
Brain Stroke	0.06 (0.00)	0.01 (0.29)	0.05 (0.00)	0.02 (0.06)
Heart Disease	0.07 (0.00)	0.02 (0.03)	0.07 (0.00)	0.03 (0.00)
Colon Cancer	0.07 (0.00)	0.02 (0.03)	0.07 (0.00)	0.02 (0.03)
Diabetes Type I	0.08 (0.00)	0.03 (0.01)	0.08 (0.00)	0.03 (0.00)
Heart AMI	0.09 (0.00)	0.04 (0.00)	0.08 (0.00)	0.05 (0.00)
Breast Cancer	0.08 (0.00)	0.03 (0.00)	0.08 (0.00)	0.03 (0.00)
Sleep Apnea	0.05 (0.00)	0.03 (0.00)	0.05 (0.00)	0.02 (0.05)
Prostate Cancer	0.08 (0.00)	0.02 (0.10)	0.08 (0.00)	0.02 (0.02)
Chronic Lyme	0.05 (0.00)	0.02 (0.02)	0.06 (0.00)	0.03 (0.01)
Ehlers-Danlos Syndrome	0.04 (0.00)	0.03 (0.01)	0.05 (0.00)	0.02 (0.02)
Alzheimers	0.06 (0.00)	0.01 (0.20)	0.07 (0.00)	0.02 (0.06)
Osteoporosis	0.04 (0.00)	0.03 (0.00)	0.05 (0.00)	0.01 (0.29)
Lupus	0.06 (0.00)	0.02 (0.02)	0.08 (0.00)	0.02 (0.08)

Pearson correlation coefficient between the observed term frequencies and the stigma words shown first. P-values are in parentheses. Values rounded to two decimal places. Those with a significant relationship (t-stat above 1.96) are shown in bold.

Table 5.7: Stigma Correlations

Disease	Stigma	Stigma and General Inquirer
Myalgic Encephalomyelitis	0.05 (0.00)	0.04 (0.00)
Depression	0.00 (0.78)	0.01 (0.47)
COPD	-0.01 (0.46)	-0.00 (0.81)
Schizophrenia	-0.01 (0.34)	0.01 (0.60)
Multiple Sclerosis	0.01 (0.47)	0.01 (0.39)
Rheumatoid Arthritis	-0.00 (0.89)	0.00 (0.96)
Lung Cancer	-0.01 (0.63)	0.00 (0.75)
Brain Stroke	-0.01 (0.30)	-0.00 (0.66)
Heart Disease	-0.00 (0.87)	0.00 (0.74)
Colon Cancer	-0.00 (0.78)	0.01 (0.31)
Diabetes Type I	-0.01 (0.26)	-0.01 (0.53)
Heart AMI	-0.01 (0.51)	-0.01 (0.59)
Breast Cancer	-0.01 (0.64)	0.01 (0.30)
Sleep Apnea	-0.01 (0.58)	-0.00 (0.79)
Prostate Cancer	0.00 (0.99)	0.00 (0.70)
Chronic Lyme	0.02 (0.02)	0.02 (0.07)
Ehlers-Danlos Syndrome	0.01 (0.28)	0.01 (0.35)
Alzheimers	-0.01 (0.62)	-0.01 (0.48)
Osteoporosis	-0.00 (0.90)	-0.00 (0.81)
Lupus	-0.01 (0.62)	-0.00 (0.98)

Pearson correlation coefficient between the observed term frequencies and the stigma words shown first. P-values are in parentheses. Values rounded to two decimal places. Those with a significant relationship are shown in bold.

Table 5.8: Regression of Word Frequencies on Stigma Words

	Stigma Words	Stigma and General Inquirer Words
ME/CFS	2.403*** (0.498)	1.559*** 0.446
constant	0.805*** (0.035)	0.808*** (0.035)
R ²	0.0023	0.0012
Adj. R ²	0.0022	0.0011
Chronic Lyme	0.912* (0.407)	0.649 . (0.364)
constant	0.567*** (0.028)	0.567*** (0.028)
R ²	0.0005	0.0003
Adj. R ²	0.0004	0.0002
EDS	0.206 0.199	0.165 0.178
constant	0.252 0.014	0.252*** 0.014
R ²	0.0001	.0001
Adj. R ²	.0000	-.0000
Footnote: Signif. codes: '***' if <0.001, '**' if <0.01, '*'if <0.05, '.' if <0.1		

5.4 Discussion and Implications

The economics literature has shown that systemic biases in judgment can lead to incorrect stereotypes concerning particular groups. These incorrect beliefs have negative welfare implications and are inefficient. Discrimination resulting from incorrect beliefs is an established issue in economics for race and gender, although searching for evidence according to disability or disease is an area that has not been developed. This paper contributes to the existing research on implicit biases in behavioral economics by being the first of which I am aware to look for differences in beliefs by disease or to use natural language processing techniques to analyze disability. I have also contributed to the body of research by designing and implementing a research strategy to test the hypothesis of differential attitudes by disease type.

The results from the previous section reveal that doctors' attitudes appear to be different for different diseases. Negative attitudes resulting in the delay or denial of treatment, or delaying investigation into symptoms, can prevent improvements in these patients' health. Patients whose disease could be kept under control can experience more severe symptoms and more severe limitations to their activities. In addition to delaying treatment, biased attitudes may lead to inefficiencies by discouraging investment in biomedical research.

Stereotyping a particular condition as mysterious could also lead to a culture of hopelessness regarding the condition in question, so deterring researchers from entering the field. It could also lead to an underestimation of the marginal benefit of funding the disease if policymakers or health researchers mistakenly believe that nothing can be done about the condition. This outcome has welfare implications for the quality of life and mortality of a population, as well as economic implications for labor market participation and GDP. Diseases like ME/CFS cost the US and UK economies billions of dollars in medical bills, care-taking costs and lost labor market productivity each year.

This analysis was able to determine with relative confidence whether there were differences in word choice between diseases. This can be reasonably assumed

to be discrimination relative to type. Type is measured along two dimensions: severity of illness and patient personality characteristics. Severity of illness can be rigorously shown and personality type can be reasonably assumed to be consistent across diseases with similar characteristics. However, this can only be shown at the aggregate level.

Limitations of this approach lie in its viability in precisely quantifying discrimination, which would require measurement of individual characteristics. To quantify incorrect beliefs or statistical discrimination (differences in behavior due to correct beliefs), there needs to be a way of measuring whether the biases maintained about the individual patients are correct. This could be done in a lab setting, which would lose the benefit of anonymous information, but would allow for more control over experimental design.

For patients with characteristics in common, discrimination could potentially be measured in terms of how many additional tests a doctor orders and whether there is more effort put forth for different conditions. This data set had 26 diseases, but if the data was collected again for more diseases, perhaps a sample of 100, then it would be possible to regress the quality of life on stigma to see if diseases with a lower quality of life experience more negative attitudes from medical professionals. Impact on quality of life could then be used as a proxy for type to see if diseases with similar impacts on quality of life face a greater degree of incorrect bias from providers.

In addition to looking at differences in attitudes by disease, another question to be investigated is whether diseases with the most stigma get less funding. As in this study, stigma could be defined with a dictionary, or it could be defined as any disease described as controversial in the data. In addition to medical discussion forums, researchers could use media and newspaper coverage to obtain text data regarding different diseases. In addition, a dictionary of medical terminology or jargon could be made to compare diseases and explore them if diseases with lower degrees of stigma are more likely to have discussions with specific medical terms. In addition, such an approach would imply fewer instances of the judgment of specific patients and more discussion of treatment.

It is also worth investigating whether high stigma or low funding is related to gender to see if underfunded diseases are more likely to be women's diseases. In ME/CFS, millions of people suffer from this disease and 80% of them are women. It is possible that it is not the only disease to reflect these trends. Are diseases described as controversial in the media more likely to be diseases that impact women? To help disentangle gender issues, researchers could look at clearly endogenously defined masculine and feminine diseases, such as gynecological diseases versus prostate cancer. For conditions that are subject to incorrect biases, the economic impact of these views could be measured - for example, in changes in how much people work before and after falling ill - for both stigmatized and non-stigmatized diseases.

Furthermore, there is no research on the impact of misdiagnosing patients with psychosomatic illnesses. If a patient is diagnosed as having a psychiatric condition and is later diagnosed with an autoimmune disease or infection that causes the same symptoms, then this is currently not tracked. Regardless of stereotypes about whole conditions, misdiagnosis of individual patients could have significant economic effects. Any research into this area should analyze the results both for gender and race to see if women or minorities experience this disproportionately.

For ME/CFS and other diseases with stigma, an analysis should be conducted into whether they are underfunded. The marginal benefit of funding is large in terms of the population size affected and disease severity. However, it is important to determine if there is under investment by looking at marginal benefits and marginal costs, which would require measuring scientific opportunity. There is some subjectivity in measuring scientific opportunity that may not be able to be completely avoided, but this analysis would still be informative.

The implications of research into implicit biases include opportunity as much as description of the current state of behavior. In situations where incorrect stereotypes are interfering with patient care, then there are chances for hospitals to introduce checklists and audits to counter the influence of heuristics. Opportunities lie with policymakers to measure the impact of diseases objectively and confirm that diseases with similar impacts and presenting similar scientific opportunities are receiving

equivalent funding. This analysis can be checked both by gender and by race to see if diseases that impact these groups are funded proportionately.

6

Conclusion

6.1 Conclusion

This paper has examined forum discussions by medical professionals to determine whether their language indicates the existence of stigma against certain medical conditions. To answer this research question, I have contributed a new dataset from a site that has not previously been used in economic research, based on a search of all available economics journals. This data included discussion by medical professionals of 26 major diseases.

If medical professionals hold biased judgments about patient groups, then it can lead to incorrect stereotypes against them and to inefficiencies both in research and treatment. These biases can be detected in language entailing real economic implications. Language is an indicator of emotions and group beliefs and behavioral economics has shown it to impact behavior and decision-making. If the attitude of the medical community impacts upon treatment, then this affects people's capacity to work and will impact in turn upon human welfare and economic productivity. The first step is to find out if there are differences in attitudes according to the category of disease, which has been my approach here.

Using novel data collected from an online discussion forum, I have used LASSO-logistic regression to determine the words that were most predictive by disease

category. I have also created a dictionary to detect biased attitudes and found that it was correlated with two of three diseases hypothesized to have stigma, while not correlated with any of the diseases that were hypothesized not to have stigma.

The results of this study show that natural language processing techniques on text data from forums can reveal doctors' attitudes toward specific medical conditions; attitudes concerning different conditions are not consistent. This relies on solidly established evidence for disease severity and an assumption that ME/CFS patients do not differ substantially in type compared to patients with MS and Lupus. ME/CFS patients are subjected to comparatively more negative attitudes than patients with diseases that share similar characteristics. This difference in attitudes concerning patient groups that share a common 'type' is evidence for medical bias based on incorrect beliefs.

These findings regarding different attitudes shows that medical decisions regarding research funding and patient care may not be objectively determined. In turn, doctors' attitudes are likely to play a role.

A

Appendix

A.1 Least Absolute Shrinkage and Selection Operator (LASSO)

The LASSO is a regression technique that aims to determine the characteristics most relevant for a given response variable. As shown in Equation A.1, it is an optimization method similar to Ordinary Least Squares Regression (OLS), which fits a line to the minimum sum of the squared residuals, but a constraint is placed on the absolute value of the parameters.

The constraint provides two benefits: regularization and feature selection. Regularization is the reduction of the complexity of the model by shrinking the coefficients. Meanwhile, feature selection is the choice of variables to keep in the model, which is done by setting the irrelevant coefficients to zero. Unlike OLS, which is a useful approach when the researcher already knows the features that are important, LASSO helps with feature selection when there are many features and it is not clear which should be in the model.

The LASSO estimate for $\hat{\beta}^{lasso}$ in the context of least squares:

$$\hat{\beta}^{lasso} = \arg \min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (\text{A.1})$$

Since the LASSO adds the absolute value of the parameters as a constraint, the variables need to be standardized to z-scores as in Equation A.2. The variables are converted to mean zero unit variance by subtracting the means and dividing by the estimated variance for each. Standardization puts all the variables on the same scale so that the coefficients are not weighted differently depending on the units in which they are measured. Once standardized, the coefficients will be comparable. In this example, the units are the same for each variable, but in terms of word frequency, the variances may not be the same so standardization is necessary (Li et al., 2017). The z-scores are defined as:

$$Z_{X_t} = \frac{X_t - \bar{X}}{\sigma} \quad (\text{A.2})$$

where:

\bar{X} is the sample mean and σ is the standard deviation (Li et al., 2017)

It is helpful to understand LASSO by first looking at the ridge regression. The ridge regression performs regularization (the shrinking of features) but not feature selection (setting the coefficients to zero). It also prevents multicollinearity. Multicollinearity is when the independent variables are nearly linear combinations of each other. One possible source of this is when there is an over-defined model, particularly when there are more observations than variables. Multicollinearity amongst the independent variables will lead to large variances in the OLS estimates, although they will remain unbiased. Ridge regressions introduce a small amount of bias in order to reduce the variance but the aim is to balance the bias-variance trade-off to produce more reliable estimates (NCSS, n.d.).

$$\begin{aligned} \text{Expected Prediction Error} &= \mathbb{E}[(Y - \hat{f}(x))^2] \\ &= \text{Bias}^2 + \text{Variance}^2 + \sigma_\epsilon^2 \end{aligned} \quad (\text{A.3})$$

where:

$\sigma_\epsilon^2 = \text{irreducible error}$

Bias = difference between estimated value and true value

Variance = spread of the data

$Y = f(x) + \epsilon$ is a function with noise

as illustrated by (Hastie et al., 2009)

The source of this bias-variance tradeoff comes from the Expected Prediction Error, or the variance of the noise, the decomposition of which is shown in Equation A.3. It describes how accurately the model can predict outcomes on new data. While the hope is that the difference between the real function and the estimate will be as small as possible, there will always be some irreducible error. There are two ways this can be minimized. On one end of the extreme, ignoring training data and merely predicting the test data with a constant would minimize the variance.

The variance of the predictions would be zero in this case because the predictions are constant. However, the bias in this case, the the amount by which the prediction is off from the real function, would be large. This oversimplification problem is called ‘underfitting’ and occurs when the model does not have enough features or when the features have been excessively regularized (Vijayakumar, 2007).

Alternatively, at the other extreme, the estimate could perfectly match the training data. In this case, the bias would be zero, but the variance would be very large. This is what is known as overfitting and occurs when the model is overly complex and captures too much noise (Vijayakumar, 2007). Generally, there is a trade-off between models that are too simple with a large bias and models that are too complex with a large variance. Usually, one can be reduced at the expense of the other as a result. In the case of high-dimensional models, there are many variables and a high risk of multicollinearity. Here, variance increases and overfitting becomes the primary concern (NCSS, n.d.).

The overfitting problem can be divided into two subsets of machine learning problems: regression and classification. In regression, the response variable is continuous, but in classification the response variable is discrete (or categorical) (Hansen, 2018b). In the application used in this study, the words were be split into disease categories to address what is then a problem of classification.

Regularization, or shrinking of the coefficients toward zero, prevents the model from being overly complex and avoids the problem of overfitting to the training

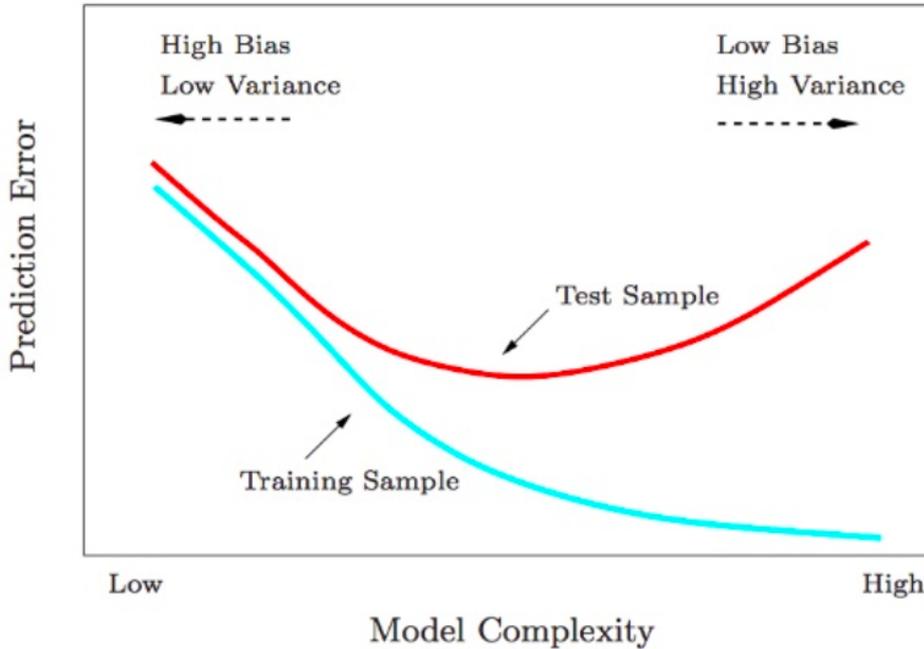


Figure A.1: Test and Training Error as a Function of Model Complexity. Overfitting occurs toward the right and underfitting toward the left on this figure. (Hastie et al., 2009)

sample. This is shown in terms of the test and training data in A.1. Regularization is done with ridge regression, which allows estimation of a model that introduces a small amount of bias in the interest of obtaining better long-term predictions.

Ridge Regression estimate for $\hat{\beta}^{ridge}$:

$$\hat{\beta}^{ridge} = \arg \min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (\text{A.4})$$

High-dimensional models are growing more frequent, to take the example of models that use thousands of genetic variants to predict disease risk or natural language models with thousands of words. Using OLS to model this situation would overfit the data excessively. It would fit the training data nearly perfectly with low bias but would not fit the testing data well and be saddled with high variance.

Ridge regression attempts to solve this problem by adding a penalty to the standard OLS problem as shown in Equation A.4. In the univariate case, it minimizes the sum of the squared residuals + $\lambda \times (\text{slope})^2$. If there is more than one

variable, as in this case, then the ridge regression would add the sum of all the parameters squared, except for the y-intercept. λ determines the size of the penalty.

When the slope of a coefficient is large, the response variable is very sensitive to changes in the independent variable. When the slope is small, the predictions are much less sensitive. The larger the penalty parameter, the more the sum of the coefficients is penalized. As λ increases, predictions for the response variable become less and less sensitive. Ridge regression reduces variance by shrinking the coefficients and making the predictions less sensitive to them. This variance was brought down by introducing a small amount of bias through the penalty parameter λ .

The intent of describing ridge regression in such detail is that a great deal can be explained about LASSO regressions by comparing them to the case of ridge regression. The key difference mathematically is that for LASSO regression, instead of squaring the sum of the slopes, it takes their absolute value. The LASSO regression is shown in A.1 and constrains the absolute value of the parameters to be less than a particular constant, which penalizes the coefficients and shrinks them toward zero. As before, the penalty contains all of the parameters save for the y-intercept¹.

When the parameters are regularized, or shrunk, they do not need to be reduced equally and less important attributes are penalized more. If the penalty parameter λ is reduced, the magnitude of the coefficients is not penalized as much and the LASSO regression approaches the linear case. When $\lambda = 0$, LASSO = OLS. On the other hand, when the size of the parameter λ is increased, the coefficients shrink. Past a certain point, all the coefficients will be exactly zero and are disregarded from the model. This elimination of irrelevant features minimizes the prediction error and is shown graphically in Figure A.1.

The advantage of LASSO over ridge regression in the high dimensional case is that ridge regression can only shrink the variables asymptotically close to zero, but LASSO can shrink them all the way to zero when the penalty parameter, λ , increases. As LASSO can exclude irrelevant variables, it is better placed to

¹y-intercept will remain as long as the response variable is not normalized, which it is not in this case as it is categorical

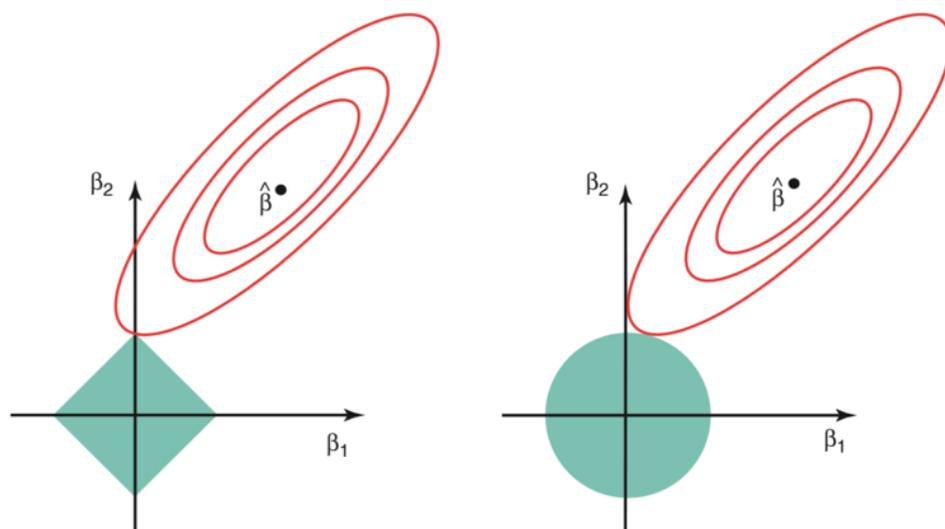


Figure A.2: Contours of error and constraint functions for LASSO (left) and ridge regression (right). The areas shaded in blue show the constraint regions $|\beta_1| + |\beta_2| \leq t$. The red ellipses are the contours of the error function $\beta_1^2 + \beta_2^2 \leq t^2$. (Hastie et al., 2009)

reduce complexity in models with many redundant or uninformative features. The resulting model is simpler and easier to interpret.

Figure A.2 contrasts the two cases. In ridge, it limits the sum of the squares of the coefficients and has a convex constraint. Due to this convexity, it tends to choose interior points, entailing that many of the coefficients are small, but not zero. In LASSO, owing to the fact that it limits the absolute value of the coefficients, it allows for corner solutions. This application of LASSO sets many of the coefficients to zero as λ is tightened. LASSO hence regularizes the features and prevents multicollinearity much like ridge, but also provides feature selection, which ridge does not.

The coefficients in the model are those that remain after the LASSO penalty is applied. There are no p-values or standard errors provided because there is no way of being certain that the selected predictors are significant in the traditional sense. The reason for this uncertainty is that the estimates from the penalized regression are biased, entailing that the standard errors are not meaningful as in OLS where the estimates are unbiased. In LASSO, it is not clear how much

the standard errors reflect the bias or the variance of the estimated coefficients (Kyung et al., 2010; Goeman et al., 2018).

The main disadvantage of using LASSO is that when features are collinear, LASSO will only select one of the features and set the other coefficient to zero. In this case, some information will be lost. For example, if the model was trying to predict housing prices and there were variables for number of rooms and square feet, these may carry similar information and only one of them would be selected in the final model. In the case of natural language processing, when there are thousands of words as potential independent variables, the loss of some information is helpful in focusing analysis. Still, it should be noted that in the case of two or more of the words providing similar information, only one of them will be selected for the final model. In addition, the model introduces bias into the model which makes interpreting standard errors difficult.

A central issue for LASSO regression is deciding the size of the penalty parameter λ . This is often done through cross-validation.

A.2 Cross-Validation

Cross Validation is a method that helps with λ selection while preventing overfitting. It tests many values for λ and uses one that results in the lowest variance. In cross-validation, the data is split into training and test samples. The training data is the set of observations used to fit the parameters. The test data is used to evaluate the model that was fit on the training data. Data is split into these two groups in order to validate the model and confirm that it picks up the general pattern in the data and does not just pick up noise.

The cross-validation method assumes the two groups are drawn from the same distribution. The test sample allows for approximation of the expected error. Removing part of the data to validate it increases the chance of underfitting because some of the underlying trends in the data may be lost. To avoid this, the data can be split into more than two groups through k-fold cross-validation (Breiman and Spector, 1992).



Figure A.3: Example of split of data for 10-fold cross-validation. This study uses 5-fold cross-validation. (Rosan, 2016)

In k-fold cross-validation, the data is randomly split into k groups of roughly equal size. The model is estimated k times and the results are averaged. Each datapoint will be in the testing set once and in the training set k-1 times. For extremely large datasets, splitting the data into only testing and training achieves good results and k-fold cross validation is not necessary. For very small datasets, N-fold cross validation is sometimes used.

Splitting data into N folds has a very low bias for expected prediction error but comes with greater computational costs in terms of memory and time because the method needs to be applied N times. As k is smaller, bias increases. For datasets with more than 200 observations, 5-fold cross-validation has a very low bias and the performance on the training set is almost the same as if the model had estimated the whole set (Hastie et al., 2009). Hence, 5 or 10-fold cross-validation is considered the standard (Breiman and Spector, 1992; Hastie et al., 2009). As there are more than 200 observations in this study, I have estimated the model using 5 folds.

The procedure of 5-fold cross-validation is:

- Randomly mix the order of the datapoints

- Split the data into 5 non-overlapping groups
- For each of the 5 groups:
 - Assign the group to be the test data
 - Assign the remaining groups to be the training data
 - Fit the model on the training data: use the set of observations in the training group to learn a model for the response variable y_i conditional on the independent variable x_i
 - Using the fitted model, generate predicted values for the y_i
 - Validate the model on the test data comparing the predicted and actual values
 - Save the prediction errors (as in A.3) and repeat for each group
- Average the errors from the last step over all 5 folds to generate the estimate of prediction error as in Figure A.3 (Hastie et al., 2009; Hansen, 2018b)

This procedure is shown visually in A.3. Each observation stays in its assigned group for the duration of the process until the groups are split again. Retaining or discarding each predictor is decided using the cross-validated accuracy or prediction error. This is also how the optimal value of λ is determined. There are generally two ways of determining λ , calculating λ_{min} or λ_{1se} . λ_{min} is the value that gives the minimum mean cross-validated error while λ_{1se} gives the value where the error is within one standard error of the minimum.

λ_{min} will not always eliminate enough predictors. When this occurs, a λ up to one standard deviation away can be used to simplify the model further. In Figure A.4 on page 92, λ_{min} is the first vertical line and λ_{1se} is the second.

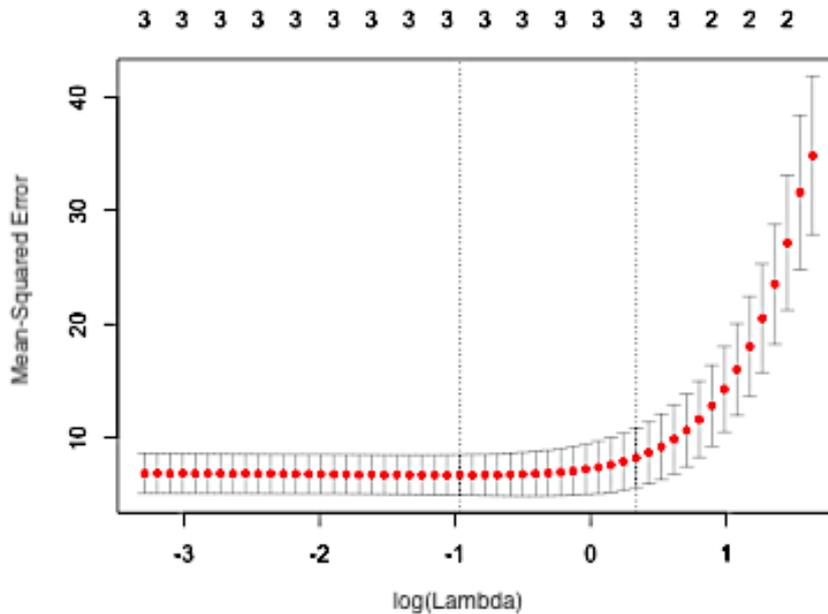


Figure A.4: Cross Validation Example. The first vertical dotted line in the figure above is λ_{min} while the second vertical line is λ_{1se} (Fonti, 2017)

A.3 Logistic Regression

Not every algorithm will work in every case. For example, linear regression cannot be used with categorical response variables. Instead, I have used logistic regression, which is a sub-type of generalized linear models. Readers familiar with logistic regressions and GLM should skip to the next subsection.

Generalized Linear Models (GLM) are a type of Linear Model where the response variable Y can have a distribution different from the normal distribution. These possible distributions include the Normal, Multinomial, Poisson or Binomial. The model in this paper will use a GLM model with a response variable with Binomial distribution and a logit link equation. Then, an upper bound on the sum of the absolute values of the model parameters will be added to make the final model a LASSO-Logistic Equation.

In a GLM model there are three elements:

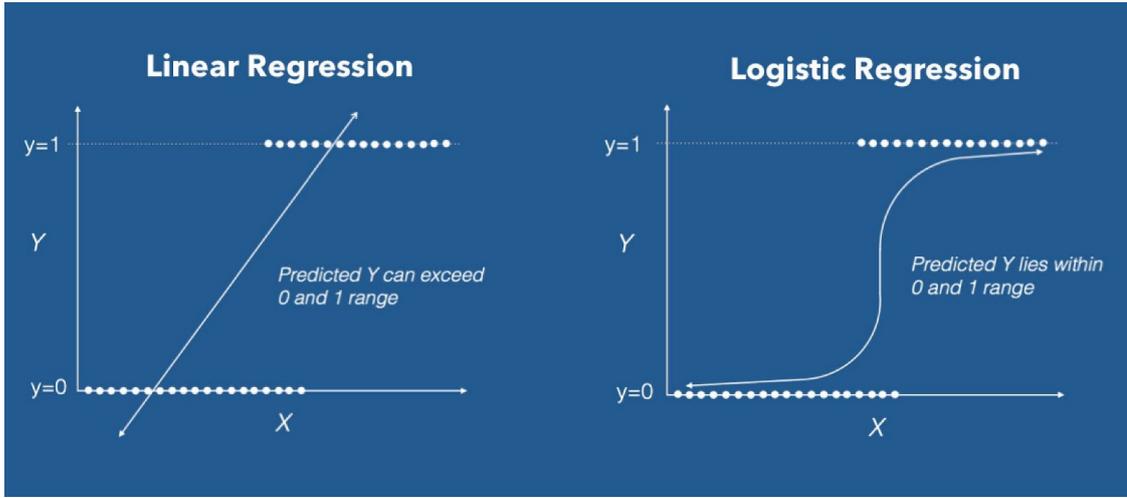


Figure A.5: Logistic regression is better for categorical variables because the logistic function restricts the range of the outcome variable to $[0,1]$ (Brid, 2018).

Random Component: probability distribution f of Y , n is the number of observations

$$Y_i \sim f \quad Y_i \in \mathbb{R}^n \quad \forall i = 1 \dots n \quad (\text{A.5})$$

Systematic Component: gives the linear combination of the variables in X , consist in a vector of predictors η_i , k is the number of variables

$$\eta_i = x_i^T \beta \quad \text{where } \beta, x_i \in \mathbb{R}^k \quad \forall i = 1 \dots k \quad (\text{A.6})$$

Link: Function g , strictly monotone, describes how the two components are related. μ is the expected value for the response variable Y_i .

$$\eta_i = g(\mu_i) \quad \forall i = 1 \dots n \quad (\text{A.7})$$

Logistic regression is a sub-type of the class of GLM models used to predict the probability of a categorical response variable. Logistic regressions produce a binary outcome variable with only two values, 0 (failure) and 1 (success), which is useful for predicting situations where a patient survives or dies, or has cancer or does not. The logistic model predicts $P(Y=1)$ as a function of the independent variable, X . In this case, the model predicts whether a given comment is about a particular disease (1) or not (0).

Random Component: The response variable Y has a *Binomial* distribution with two parameters, n_i , the number of observations, and π_i , the probabilities.

$$Y_i \sim \text{Binomial}(n_i, \pi_i) \quad Y_i \in \mathbb{R}^n \quad \forall i = 1 \dots n \quad (\text{A.8})$$

y_i is a realization of the random variable Y_i while the probability distribution function of Y_i is:

$$\text{Pr}(y_i = Y_i) = \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \quad (\text{A.9})$$

Systematic Component: Here is *logit* function, or log-odds.

$$\eta_i = x_i^T \beta \quad \text{where } \beta, x_i \in \mathbb{R}^k \quad \forall i = 1 \dots k \quad (\text{A.10})$$

Link: Function g , strictly monotone, describes how the two components are related. μ is the expected value for the response variable Y_i .

$$\eta_i = \text{logit}(\pi) = \log\left(\frac{\pi}{1 - \pi}\right) = \sum_{j=0}^k \beta_j x^j \quad (\text{A.11})$$

If a linear regression were to be used, then it would pose a problem because the outcome is binary between 0 or 1, but the predictor β, x_i can take any value in the real numbers. Linear models would predict values outside the required range (0 to 1). The logistic regression creates a logistic curve which is restricted to between the values 0 and 1. To do this regression, first the probabilities are transformed into odds so 1/3 would be 1 to 2. Then, taking logs allows the probabilities (all between 0 and 1) to be mapped to the entire range.

The above equations constitute a GLM model with a binary response variable and a logit link equation. The regression coefficients β_j give the change in the *logit* probability associated with a change in the j -th feature.

The probabilities in this case are:

$$P(Y_i = 1|x_i) = \pi_i(x_i^T \beta) = \frac{\exp\{x_i^T \beta\}}{1 + \exp\{x_i^T \beta\}}, \quad 1 \leq i \leq n \quad (\text{A.12})$$

The logistic regression optimizes the sum of the likelihoods rather than the sum of squared residuals because logistic regression is solved using maximum likelihood.

It can also be modified to examine cases when the dependent variable has more than two categories which are unordered. This is called multinomial logistic regression and is described later when it is used. The binary model is used to determine whether a given comment is ME/CFS (1) or not (0). The multinomial regression will be used to distinguish between multiple disease categories.

A.4 LASSO-Logistic Regression

A LASSO-Logistic Regression makes one more modification to the previous equations by putting an upper limit on the sum of the absolute values of the model parameters. The negative log-likelihood is penalized with an l_1 -norm (Rodriguez, 2007). Where n is the number of observations and p the number of columns in X , the negative log-likelihood is:

$$l_n(\beta) = - \sum_{i=1}^n \log(P_\beta(Y_i|X_i)) = \sum_{i=1}^n \left\{ -Y_i \left(\sum_{j=0}^p \beta_j x^j \right) + \log \left(1 + \exp \left(\sum_{j=0}^p \beta_j x^j \right) \right) \right\} \quad (\text{A.13})$$

Y_1, \dots, Y_n are independent binary response variables and x_1, \dots, x_n are the covariates that correspond. $x_i = (1, x_{i1}, \dots, x_{ik})^T$ for $1 \leq i \leq n$. $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ is a $(p + 1)$ -dimensional vector.

In terms of the loss function ρ :

$$\rho(\beta) = -y \left(\sum_{j=0}^p \beta_j x^j \right) + \log \left(1 + \exp \left(\sum_{j=0}^p \beta_j x^j \right) \right) \quad (\text{A.14})$$

The LASSO estimator for this is:

$$\hat{\beta}(\lambda) = \arg \min_{\beta} \left(n^{-1} \sum_{i=1}^n \rho_\beta(X_i, Y_i) + \lambda \|\beta\|_1 \right) \quad (\text{A.15})$$

and:

$$\|\beta\|_1 = \sum_{j=1}^p |\beta^j| \quad (\text{A.16})$$

Bibliography

- Action for M.E. (2015). Close to collapse: An interim report on access to social care and advocacy for people with ME/CFS, *Technical report*, Action for ME.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002). How Cells Obtain Energy from Food, *Molecular Biology of the Cell. 4th edition* .
- Allan, N. (2013). Who Will Tomorrow’s Historians Consider Today’s Greatest Inventors?, *The Atlantic* .
- Antweiler, W. and Frank, M. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards - Antweiler - 2004 - The Journal of Finance - Wiley Online Library, *The Journal of Finance* **59**(3).
- Archive, A. (2019). 'Medical Sexism': Women’s heart disease symptoms often dismissed.
- Balogh, E. P., Miller, B. T., Ball, J. R., Care, C. o. D. E. i. H., Services, B. o. H. C., Medicine, I. o. and The National Academies of Sciences, E. (2015). *Improving Diagnosis in Health Care: Overview of Diagnostic Error in Health Care*, National Academies Press (US).
URL: *ec*
- Bandiera, O., Hansen, S., Prat, A. and Sadun, R. (2017). CEO Behavior and Firm Performance, *Technical Report w23248*, National Bureau of Economic Research, Cambridge, MA.
- Bartelme, T. (2005). Luck or Skill? A-Team Players Don’t Get Sued, *Physician Executive; Tampa* **31**(5): 22–6.
- Belloni, A., Chernozhukov, V. and Hansen, C. (2011). Inference for High-Dimensional Sparse Econometric Models, *arXiv* . arXiv: 1201.0220.
- Bergmanis, T. and Goldwater, S. (2018). Context Sensitive Neural Lemmatization with Lematus, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Association for Computational Linguistics, New Orleans, Louisiana, pp. 1391–1400.
- Berner, E. S. and Graber, M. L. (2008). Overconfidence as a Cause of Diagnostic Error in Medicine, *The American Journal of Medicine* **121**(5): S2–S23.
- Bertrand, M., Chugh, D. and Mullainathan, S. (2005). Implicit Discrimination, *American Economic Review* **95**(2): 94–98.
- Bertrand, M. and Duflo, E. (2016). Field Experiments on Discrimination, p. w22014.

- Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination, *The American Economic Review* **94**(4): 991–1013.
- Blease, C., Carel, H. and Geraghty, K. (2017). Epistemic injustice in healthcare encounters: evidence from chronic fatigue syndrome, *Journal of Medical Ethics* **43**(8): 549–557.
- Blei, D. M. (2003). Latent Dirichlet Allocation, *Journal of Machine Learning Research* **3**: 30.
- Bohren, A., Imas, A. and Rosenberg, M. (2018a). The Language of Discrimination: Using Experimental versus Observational Data, *AEA Papers and Proceedings* **108**: 169–74.
- Bohren, J. A., Imas, A. and Rosenberg, M. (2018b). The Dynamics of Discrimination: Theory and Evidence, *Penn Institute for Economic Research* p. 76.
- Bollen, J., Mao, H. and Zeng, X.-J. (2011). Twitter mood predicts the stock market, *Journal of Computational Science* **2**(1): 1–8.
- Bordalo, P., Coffman, K., Gennaioli, N. and Shleifer, A. (2016a). Beliefs about Gender, *Working Paper*. .
- Bordalo, P., Coffman, K., Gennaioli, N. and Shleifer, A. (2016b). Stereotypes, *The Quarterly Journal of Economics* **131**(4): 1753–1794.
- Brea, J. (2017). Millions of Women Suffer — But Doctors Still Don't Take It Seriously.
- Breiman, L. and Spector, P. (1992). Submodel Selection and Evaluation in Regression. The X-Random Case, *International Statistical Review* **60**(3): 291–319.
- Brid, R. S. (2018). Logistic Regression.
- Bureau of Labor Statistics (2018). Labor Force Statistics from the Current Population Survey.
- Burgess, G. (2018). Interview with Jane Colby, Executive Director of Tymes Trust Charity (The Young ME Sufferers).
- Burke, J. (2016). Neuroimaging overuse is more common in Medicare compared with the VA, *Neurology* **87**(8).
- Butler, C. C., Evans, M. and Greaves, D. (2004). Medically Unexplained Symptoms: The Biopsychosocial Model Found Wanting, *Journal of the Royal Society of Medicine* **97**: 4.
- Carruthers et al. (2011). Myalgic encephalomyelitis: International Consensus Criteria, *Journal of Internal Medicine* **270**: 327–338.
- CDC (2018a). Treating the Most Disruptive Symptoms First and Preventing Worsening of Symptoms | Clinical Care of Patients | Healthcare Providers | Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) | CDC.
- CDC (2018b). What is ME/CFS?

- Chen, M. K. (2013). The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets, *American Economic Review* **103**(2): 690–731.
- Chiaromonte, G. R. (2007). Physicians' Gender Bias in the Diagnosis, Treatment, and Interpretation of Coronary Heart Disease Symptoms, *Stony Brook University* .
- Chinco, A., Clark-Joseph, A. D. and Ye, M. (2018). Sparse Signals in the Cross-Section of Returns, *The Journal of Finance* **74**(1): 449–492.
- Chugh, D. (2004). Societal and Managerial Implications of Implicit Social Cognition: Why Milliseconds Matter, *Social Justice Research* **17**(2): 203–222.
- Cohen, J. (2015). NIH refocuses research into chronic fatigue syndrome.
- Colby, J. (2014). False Allegations of Child Abuse in Cases of Childhood Myalgic Encephalomyelitis (ME), *Argument & Critique* .
- Davis, M. (2017). Is ME/CFS an autoimmune disease?
- Davis, R. (2016). Severely Ill Big Data Study.
- Devasahayam, A., Lawn, T., Murphy, M. and White, P. D. (2012). Alternative diagnoses to chronic fatigue syndrome in referrals to a specialist service: service evaluation survey, *JRSM short reports* **3**(1): 4.
- Dimmock, M., Mirin, A. and Jason, L. (2016). Estimating the disease burden of ME/CFS in the United States and its relation to research funding, *Journal of Medicine and Therapeutics* **1**(1).
- Dr. Nigel Speight (2014). Wetenschap voor Patiënten (Science to patients) Seminar 28: Interview with Dr. Nigel Speight.
URL: <https://youtu.be/XcRZo1vO53c>
- Dusenbery, M. (2017). *Doing Harm*, 1 edn, HarperOne.
- Eisenstein, J., Smith, N. A. and Xing, E. P. (2011). Discovering sociolinguistic associations with structured sparsity, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* pp. 1365–1374.
- Engelberg, J. E. and Parsons, C. A. (2011). The Causal Impact of Media in Financial Markets, *The Journal of Finance* **66**(1): 67–97.
- Ewens, M., Tomlin, B. and Wang, L. C. (2014). Statistical Discrimination or Prejudice? A Large Sample Field Experiment, *Review of Economics and Statistics* **96**(1): 119–134.
- Falk Hvidberg, M., Brinth, L. S., Olesen, A. V., Petersen, K. D. and Ehlers, L. (2015). The Health-Related Quality of Life for Patients with Myalgic Encephalomyelitis / Chronic Fatigue Syndrome (ME/CFS), *PLOS ONE* **10**(7): e0132421.
- Faulkner, G. (2016). In the Expectation of Recovery: Misleading Medical Research and Welfare Reform, *The Centre for Welfare Reform* p. 64.

- Fonti, V. (2017). Feature Selection using LASSO, *VU Amsterdam* .
- Foundation, N. (2019). The Nobel Prize in Physiology or Medicine 1962.
- Franco, M. (2017). The Beginner’s Guide to Reddit.
- Gentzkow, M., Kelly, B. and Taddy, M. (2017). Text as Data.
- Gentzkow, M. and Shapiro, J. M. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers, *Econometrica* **78**(1): 35–71.
- Gentzkow, M., Shapiro, J. M. and Taddy, M. (2016). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech, *Working Paper 22423*, National Bureau of Economic Research.
- Geraghty, K. J. and Blease, C. (2018). Myalgic encephalomyelitis/chronic fatigue syndrome and the biopsychosocial model: a review of patient harm and distress in the medical encounter, *Disability and Rehabilitation* pp. 1–10.
- Geraghty, K. J. and Esmail, A. (2016). Chronic fatigue syndrome: is the biopsychosocial model responsible for patient dissatisfaction and harm?, *The British journal of general practice : the journal of the Royal College of General Practitioners* **66**(649): 437–8.
- Ghaemi, S. N. (2010). *The Rise and Fall of the Biopsychosocial Model: Reconciling Art and Science in Psychiatry*, JHU Press.
- Ghaemi, S. N. (2011). The Biopsychosocial Model in Psychiatry: A Critique, *Existenz: An International Journal in Philosophy, Religion, Politics, and the Arts* **6**(1): 8.
- Glover, D., Pallais, A. and Pariente, W. (2017). Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores, *Quarterly Journal of Economics* **132**(3): 1219–1260.
- Goeman, J., Meijer, R. and Chaturvedi, N. (2018). L1 and L2 Penalized Regression Models.
URL: <https://cran.r-project.org/web/packages/penalized/vignettes/penalized.pdf>
- Grant, J. and Buxton, M. J. (2018). Economic returns to medical research funding, *BMJ Open* **8**(9).
- Hansen, S. (2018a). Machine Learning Methods for Economists: Introduction.
- Hansen, S. (2018b). Machine Learning Methods for Economists: Probability Models for Discrete Data.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2 edn, Springer.
- Hayward, R. A. (2002). Counting Deaths Due to Medical Errors, *JAMA* **288**(19): 2404–2404.
- Heyhoe, J., Birks, Y., Harrison, R., O’Hara, J. K., Cracknell, A. and Lawton, R. (2016). The role of emotion in patient safety: Are we brave enough to scratch beneath the surface?, *Journal of the Royal Society of Medicine* **109**(2): 52–58.

- Iedema, R., Jorm, C. and Lum, M. (2009). Affect is central to patient safety: the horror stories of young anaesthetists, *Social Science & Medicine (1982)* **69**(12): 1750–1756.
- Inquirer, G. (2002). Inquirer – Use and Comparisons.
- Internet, A. (2019). Reddit.com Traffic, Demographics and Competitors - Alexa.
- IOM (2012). *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*, Institute of Medicine.
- IOM (2015a). *Beyond Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Redefining an Illness*, National Academies Press (US), Washington (DC).
- IOM (2015b). Report Brief: Redefining an Illness, *Technical report*, Institute of Medicine of the National Academies.
- Jackson, C. (1995). Southern Regional Benefits: Chronic Fatigue Syndrome Management Program, *Technical report*, Unum Life Insurance Company.
- Jason, L. A., Ohanian, D., Brown, A., Sunnquist, M., McManimen, S., Klebek, L., Fox, P. and Sorenson, M. (2017). Differentiating Multiple Sclerosis from Myalgic Encephalomyelitis and Chronic Fatigue Syndrome, *Insights in Biomedicine* **2**(2).
- Jason, L. A., Ropacki, M. T., Santoro, N. B., Richman, J. A., Heatherly, W., Taylor, R., Ferrari, J. R., Haneydavis, T. M., Rademaker, A., Dupuis, J., Golding, J., Plioplys, A. V. and Plioplys, S. (2011). A Screening Instrument for Chronic Fatigue Syndrome, *Journal of Chronic Fatigue Syndrome* **3**(1): 39–59.
- Kennedy, M. (2016). Women and low-income patients 'twice as likely to suffer delays in brain tumour diagnosis', *The Guardian* .
- Kindlon, T. and Me, I. (2011). Reporting of Harms Associated with Graded Exercise Therapy and Cognitive Behavioural Therapy in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome, *Bulletin of the IACFS/ME* **19**: 59–111.
- Klimas, N. G., Salvato, F. R., Morgan, R. and Fletcher, M. A. (1990). Immunologic Abnormalities in Chronic Fatigue Syndrome, *Journal of Clinical Microbiology* **28**: 8.
- Komaroff, A. L., Fagioli, L. R., Doolittle, T. H., Gandek, B., Gleit, M. A., Guerriero, R. T., Kornish, R. J., Ware, N. C., Ware, J. E. and Bates, D. W. (1996). Health status in patients with chronic fatigue syndrome and in general population and disease comparison groups, *The American Journal of Medicine* **101**(3): 281–290.
- Kyung, M., Gill, J., Ghosh, M. and Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos, *Bayesian Analysis* **5**(2): 369–411.
- Leeper, T. J. (2018). Interpreting Regression Results using Average Marginal Effects with R's margins, p. 32.
- Li, T., Jing, B., Ying, N. and Yu, X. (2017). Adaptive Scaling, *Annals of Applied Statistics* .

- Lichtman Judith H., Leifheit-Limson Erica C., Watanabe Emi, Allen Norrina B., Garavalia Brian, Garavalia Linda S., Spertus John A., Krumholz Harlan M. and Curry Leslie A. (2015). Symptom Recognition and Healthcare Experiences of Young Women With Acute Myocardial Infarction, *Circulation: Cardiovascular Quality and Outcomes* **8**(2_suppl_1): S31–S38.
- Lin, J.-M. S., Resch, S. C., Brimmer, D. J., Johnson, A., Kennedy, S., Burstein, N. and Simon, C. J. (2011). The economic impact of chronic fatigue syndrome in Georgia: direct and indirect costs, *Cost Effectiveness and Resource Allocation : C/E* **9**: 1.
- List, J. (2004). The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field, *The Quarterly Journal of Economics* .
- Lucca, D. O. and Trebbi, F. (2009). Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements, *National Bureau of Economic Research* .
- Maes, M. and Twisk, F. N. (2010). Chronic fatigue syndrome: Harvey and Wessely’s (bio)psychosocial model versus a bio(psychosocial) model based on inflammatory and oxidative and nitrosative stress pathways, *BMC Medicine* **8**(1): 35.
- Manning, C. (2008). Stemming and lemmatization, *Introduction to Information Retrieval*, Cambridge University Press.
- Monaghan, C. (2019). Question to Theresa May Regarding Child Protection Proceedings.
- Montoya, J. (n.d.). Dr. Montoya - Post screening discussion.
URL: <http://voicesfromtheshadowsfilm.co.uk/mill-valley-discussion/>
- Montoya, J. G., Holmes, T. H., Anderson, J. N., Maecker, H. T., Rosenberg-Hasson, Y., Valencia, I. J., Chu, L., Younger, J. W., Tato, C. M. and Davis, M. M. (2017). Cytokine signature associated with disease severity in chronic fatigue syndrome patients, *Proceedings of the National Academy of Sciences* **114**(34): E7150–E7158.
- Mosca Lori, Linfante Allison H., Benjamin Emelia J., Berra Kathy, Hayes Sharonne N., Walsh Brian W., Fabunmi Rosalind P., Kwan Johnny, Mills Thomas and Simpson Susan Lee (2005). National Study of Physician Awareness and Adherence to Cardiovascular Disease Prevention Guidelines, *Circulation* **111**(4): 499–510.
- Nakatomi, Y., Mizuno, K., Ishii, A., Wada, Y., Tanaka, M., Tazawa, S., Onoe, K., Fukuda, S., Kawabe, J., Takahashi, K., Kataoka, Y., Shiomi, S., Yamaguti, K., Inaba, M., Kuratsune, H. and Watanabe, Y. (2014). Neuroinflammation in Patients with Chronic Fatigue Syndrome/Myalgic Encephalomyelitis: An 11c-(R)-PK11195 PET Study, *Journal of Nuclear Medicine* **55**(6): 945–950.
- Naviaux, R. K., Naviaux, J. C., Li, K., Bright, A. T., Alaynick, W. A., Wang, L., Baxter, A., Nathan, N., Anderson, W. and Gordon, E. (2017). Metabolic features of chronic fatigue syndrome, *Proceedings of the National Academy of Sciences (PNAS)* **114**(18): 10.
- NCSS (n.d.). Ridge Regression.

- Newman-Toker, D. E., Moy, E., Valente, E., Coffey, R. and Hines, A. L. (2014). Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample, *Diagnosis (Berlin, Germany)* **1**(2): 155–166.
- Nguyen, T., Johnston, S., Clarke, L., Smith, P., Staines, D. and Marshall-Gradisnik, S. (2017). Impaired calcium mobilization in natural killer cells from chronic fatigue syndrome/myalgic encephalomyelitis patients is associated with transient receptor potential melastatin 3 ion channels: Calcium ion mobilization in natural killer cells, *Clinical & Experimental Immunology* **187**(2): 284–293.
- Nicol, W. (2018). What is Reddit? A Beginner's Guide to the Front Page of the Internet.
- NIH (2016). Report on NIH Funding vs. Global Burden of Disease -NIH Research Portfolio Online Reporting Tools (RePORT).
- Nyquist, J. G. (2014). What Doctors Feel: How Emotions Affect the Practice of Medicine, *The Journal of Chiropractic Education* **28**(2): 173–174.
- NYSDOPH (2018). Myalgic encephalomyelitis/chronic fatigue syndrome.
- of Commons, H. (2018). PACE Trial: People with ME, *Technical Report 636*, UK Parliament - House of Commons, Westminster Hall.
URL: <https://hansard.parliament.uk/commons/2018-02-20/debates/990746c7-9010-4566-940d-249f5026ff73/pacetrialleoplewithme>
- of Commons, H. (2019). Appropriate ME Treatment, *Technical Report 653*, UK Parliament - House of Commons, Westminster Hall.
URL: <https://hansard.parliament.uk/commons/2019-01-24/debates/FA1BBC27-37A7-4BFD-A2C0-A58B57F41D4D/AppropriateMETreatment>
- Office for National Statistics (2019). Labour market overview, UK: March 2019.
- Ohlheiser, A. (2017). Reddit will limit the reach of a pro-Trump board and crack down on its 'most toxic users', *Washington Post* .
- OMF (2019). End ME/CFS Project.
- Ondrich, J., Ross, S. and Yinger, J. (2003). Now You See It, Now You Don't: Why Do Real Estate Agents Withhold Available Houses from Black Customers?, *Syracuse University Center for Policy Research* .
- Optimum Health Clinic Foundation (2017). Counting The Cost, *Technical report*, Optimum Health Clinic Foundation.
- Pendergrast, T., Brown, A., Sunnquist, M., Jantke, R., Newton, J. L., Strand, E. B. and Jason, L. A. (2016). Housebound versus nonhousebound patients with myalgic encephalomyelitis and chronic fatigue syndrome, *Chronic Illness* **12**(4): 292–307.
- Pope, J. H., Feldman, J. A. and Selker, H. P. (2000). Missed Diagnoses of Acute Cardiac Ischemia in the Emergency Department, *The New England Journal of Medicine* p. 8.
- Popper, K. (1989). Two meanings of falsifiability, *Dictionary of epistemology* .

- Rally (2012). Agile Metrics - Modeling, Simulation, and Data Mining.
- Rehmyer, J. (2015). A Disease Doctors Refuse to See, *New York Times* .
- Rehmyer, J. and Tuller, D. (2017). Getting It Wrong on Chronic Fatigue Syndrome, *New York Times* .
- r/medicine* (2019).
URL: <https://www.reddit.com/r/medicine/>
- Rodriguez (2007). Logit Models for Binary Data, *Technical report*, Princeton University.
URL: <https://data.princeton.edu/wws509/notes/c3.pdf>
- Rosan, K. (2016). ML Learning Log.
- Rowe, P. C., Underhill, R. A., Friedman, K. J., Gurwitt, A., Medow, M. S., Schwartz, M. S., Speight, N., Stewart, J. M., Vallings, R. and Rowe, K. S. (2017). Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Diagnosis and Management in Young People: A Primer, *Frontiers in Pediatrics* **5**.
- Roxanne Dryden-Edwards, MD (n.d.). Morgellons Disease Treatment, Symptoms & Causes.
- Sarsons, H. (2017a). Gender Differences in Recognition for Group Work.
- Sarsons, H. (2017b). Interpreting Signals in the Labor Market: Evidence from Medical Referrals, p. 71.
- Sarsons, H. and Xu, G. (2015). Condence Men? Gender and Condence: Evidence among Top Economists, p. 26.
- Schwartz, H. A. (2013). Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach, *PLOS ONE* .
- Science, A. A. f. t. A. o. (2001). A History of the Human Genome Project, *Science* **291**(5507): 1195–1195.
- Scott, S. L. and Varian, H. R. (2015). Bayesian Variable Selection for Nowcasting Economic Time Series, *Economic Analysis of the Digital Economy* pp. 119–135.
- Shakespeare, T., Watson, N. and Alghaib, O. A. (2017). Blaming the victim, all over again: Waddell and Aylward’s biopsychosocial (BPS) model of disability, *Critical Social Policy* **37**(1): 22–41.
- Sussex, J., Feng, Y., Mestre-Ferrandiz, J., Pistollato, M., Hafner, M., BurrIDGE, P. and Grant, J. (2016). Quantifying the economic impact of government and charity funding of medical research on private research and development funding in the United Kingdom, *BMC Medicine* **14**(1).
- Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *The Journal of Finance* **62**(3): 1139–1168.
- Tucker, M. (2015). IOM Gives Chronic Fatigue Syndrome a New Name and Definition.

- Tucker, M. (2019). New Findings Elucidate Potentially Treatable Aspects of ME/CFS.
- Twisk, F. N. M. and Maes, M. (2009). A review on cognitive behavioral therapy (CBT) and graded exercise therapy (GET) in myalgic encephalomyelitis (ME) / chronic fatigue syndrome (CFS): CBT/GET is not only ineffective and not evidence-based, but also potentially harmful for many patients with ME/CFS, *Neuro Endocrinology Letters* **30**(3): 284–299.
- Union, O. a. E. (2016). Health at a Glance: Europe 2016, *9789264265592*, OECD and European Union.
- Valdez, A. R., Hancock, E. E., Adebayo, S., Kiernicki, D. J., Proskauer, D., Attewell, J. R., Bateman, L., DeMaria, A. J., Lapp, C. W., Rowe, P. C. and Proskauer, C. (2019). Estimating Prevalence, Demographics, and Costs of ME/CFS Using Large Scale Medical Claims Data and Machine Learning, *Frontiers in Pediatrics* **6**.
- Vijayakumar, S. (2007). The Bias-Variance Tradeoff, *University Edinburgh* .
- Vink, M. and Vink-Niese, A. (2018). Graded exercise therapy for myalgic encephalomyelitis/chronic fatigue syndrome is not effective and unsafe. Re-analysis of a Cochrane review, *Health Psychology Open* **5**(2).
- Wilshire, C. E., Kindlon, T., Courtney, R., Matthees, A., Tuller, D., Geraghty, K. and Levin, B. (2018). Rethinking the treatment of chronic fatigue syndrome—a reanalysis and evaluation of findings from a recent major trial of graded exercise and CBT, *BMC Psychology* **6**(1).
- Wilshire, C., Kindlon, T., Matthees, A. and McGrath, S. (2017). Can patients with chronic fatigue syndrome really recover after graded exercise or cognitive behavioural therapy? A critical commentary and preliminary re-analysis of the PACE trial, *Fatigue: Biomedicine, Health & Behavior* **5**(1): 43–56.
- Wormser, G. P. and Shapiro, E. D. (2009). Implications of Gender in Chronic Lyme Disease, *Journal of Women's Health* **18**(6): 831–834.
- Wright, N. (2018). Time for Unrest: Why patients with ME are demanding justice, *The Independent* .
- Wu, A. (2017). Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum, *IDEAS Working Paper Series from RePEc, 2017* .