Doctors' attitudes toward specific medical conditions ☆

Brooke Scoles [a,1], Catia Nicodemo [b,*]

[a] London School of Economics, Houghton St, Hol-born, London WC2A 2AE, United Kingdom
[b] University of Oxford, Department of Primary Care, and University of Verona Italy

ABSTRACT

This study uses machine learning and natural language processing tools to examine the language used by healthcare professionals on a global online forum. It contributes to an underdeveloped area of knowledge, that of physician attitudes toward their patients. Using comments left by physicians on Reddit's "Medicine" subreddit (r/medicine), we test if the language from online discussions can reveal doctors' attitudes toward specific medical conditions. We focus on a set of chronic conditions that usually are more stigmatized and compare them to ones well accepted by the medical community. We discovered that when comparing diseases with similar traits, doctors discussed some conditions with more negative attitudes. These results show bias does not occur only along the dimensions traditionally analyzed in the economics literature of gender and race, but also along the dimension of disease type. This is meaningful because the emotions associated with beliefs impact physicians' decision making, prescribing behavior, and quality of care. First, we run a binomial LASSO-logistic regression to compare a range of 21 diseases against myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), depression, and the autoimmune diseases multiple sclerosis and rheumatoid arthritis. Next, we use dictionary methods to compare five more chronic diseases: Lyme disease, Ehlers-Danlos syndrome (EDS), Alzheimer's disease, osteoporosis, and lupus. The results show physicians discuss ME/CFS, depression, and Lyme disease with more negative language than the other diseases in the set. The results for ME/CFS included over four times more negative words than the results for depression.

1. Introduction

Prejudice or stereotyped beliefs against people with certain health conditions have real health and economic impacts. Systemic biases in judgment of patients, if they exist, can lead to incorrect stereotypes about that patient group and influ-

ence their care. The field of economics has developed literature on stereotyping for gender and race. This paper provides evidence for another dimension of bias, that of medical providers toward patients with specific disease types.

Studies by Sarsons (2017), Bertrand et al. (2005), and Bertrand and Duflo (2016) show beliefs influence behavior and affect real economic outcomes. Economics researchers can use natural language processing (NLP) and machine learning (ML) techniques to measure these beliefs in language (Chen, 2013; Bohren et al., 2018; Gentzkow et al., 2019; Lucca and Trebbi, 2009). Our research applies this to a health context, to learn whether doctors have biased beliefs toward their patients. It is necessary to understand physicians' perspectives, beliefs, and emotions because their beliefs impact their healthcare decision-making and ultimately healthcare markets. Goddu et al. (2018) show bias in patient notes transmits between physicians and affects prescribing decisions. If patients in stigmatized groups receive a lower quality of care, they may be more likely to remain ill and suffer more economic consequences.

In this study, we answer the following question: do medical professionals judge chronically ill patients differently depending on their specific disease? Patients deserve equal and compassionate treatment. If healthcare providers take preconceived prejudices into practice, changes to training and recruitment may affect the quality of care. These prejudices may distort the quantification of the number of people who suffer from these diseases and may distort research priorities if diseases the medical community considers undeserving receive disproportionately low attention compared to their severity and prevalence. Thus, a deeper understanding of this topic is important to ensure compassionate patient treatment, to shape healthcare policies, and to promote efficient functioning of healthcare markets.

Following the approach of Wu (2017), who studied beliefs toward gender on the economic job market forum, we use an online forum for physicians and other medical professionals. For this study we use a novel database from Reddit's "Medicine" subreddit (r/medicine). Using the Application Programming Interface (API)[2] for the site, we requested the comments from 2008 until 2019, and used them to test if the language from online discussions reveals doctors' attitudes toward specific medical conditions. Users post anonymously, meaning language may be less edited than in other places, revealing physicians' true opinions.

As noted by Wu (2017), findings from lab experiments have limitations. For instance, in a laboratory experiment, people know they are observed, and may not reveal their true beliefs. Moreover, not all biases are easily observable in behavior. An alternative approach is to study people's behavior on anonymous online forums. Measuring stereotypes in this setting is potentially more accurate because there is less incentive to edit behavior to meet social expectations (Gentzkow et al., 2019; Wu, 2017). Unlike comments made publicly, social media data is rawer and richer. It is possible to gather information and signals that are not available in interviews or lab environments where doctors may self-censor.

We consider the language toward three groups of chronic illnesses: myalgic encephalomyelitis/ chronic fatigue syndrome (ME/CFS), depression, and autoimmune disease. ME/CFS is a long-term illness that affects many body systems. Usually individuals with ME/CFS are often not able to do their usual activities. People with ME/CFS have severe fatigue, sleep, thinking and concentrating problems, pain, and dizziness. We included depression to compare ME/CFS to another potentially stigmatized condition and to reveal differences in language between ME/CFS and mental illness. The autoimmune disease category includes multiple sclerosis (MS) and rheumatoid arthritis (RA). We analyzed MS and RA because they have similar characteristics to ME/CFS, including comparable disease type, symptoms, and demographics. For each of the three categories we compared results against the range of 21 diseases from (Hvidberg et al., 2015) in Fig. 3.

In Section 5 we use dictionary methods to compare five additional chronic diseases: Lyme disease, Ehlers-Danlos syndrome (EDS), Alzheimer's disease, osteoporosis, and lupus. While the last three diseases are very well accepted and recognized by doctors and society, the first three are usually more stigmatized. We chose these diseases because they are a mix of both poorly understood and established diseases. Our objective is to study perceptions of "invisible illness" that affect many people but are often stigmatized (Pilkington et al., 2020).

We use a LASSO-logistic regression to compare ME/CFS to other major diseases and to test for differences in attitudes between conditions. LASSO is a machine-learning tool used to choose relevant features from high-dimensional datasets. We used this approach because text data is inherently high-dimensional. In text data, there are thousands of potential variables, and it is not obvious which are important to include in the model. In addition to determining which words are relevant, LASSO can determine the size of the coefficients for each word (Fonti, 2017).[3]

Our regression results show that medical opinion may not be objectively determined but determined partly by prejudice. Moreover, the comments vary between diseases, showing that doctors' attitudes are not consistent. Attitudes toward patients differ depending on the disease. This result holds when comparing diseases that have defining characteristics in common.

To further this analysis, we classified words into a dictionary to measure the correlation between the word-frequencies for the diseases in the dataset and the stigmatizing words. The framework for identifying these words is discussed in Section 5. We found that the stigma words were correlated with ME/CFS but not for the other diseases. Then, using five additional diseases, we measured the correlations again, hypothesizing that we would find stigma in two of these five conditions. Our hypothesis was correct for four of the five additional diseases, showing that the words classified as stigmatized may generalize to other conditions. The results of each level of analysis were similar: each part showed that doctors used

---

[2] An Application Programming Interface (API), is a set of functions that allows a developer to access the data of an operating system or application. APIs are the primary means of external data access to social media data (Hogan, 2018).

[3] In the economics literature, 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).

Disease Prevalence and Research Funding per Patient (in US)
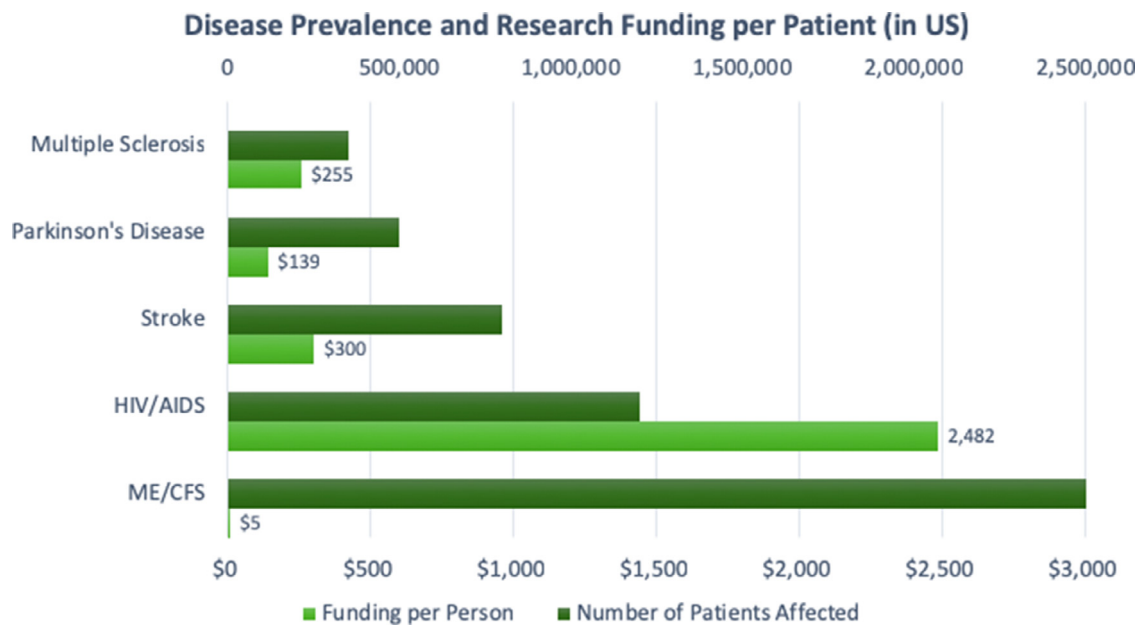
Fig. 1. In the US, ME/CFS affects more people than many major diseases but receives lower funding per patient. Funding is $5 per patient per year. Diseases above were chosen because they all have a severe impact on quality of life. Data from Dimmock et al. (2016); American College of Rheumatology, 2016; NIH, 2016.

Funding vs US Burden of Disease - 2013

$$y = 107326x^{0.5541}$$
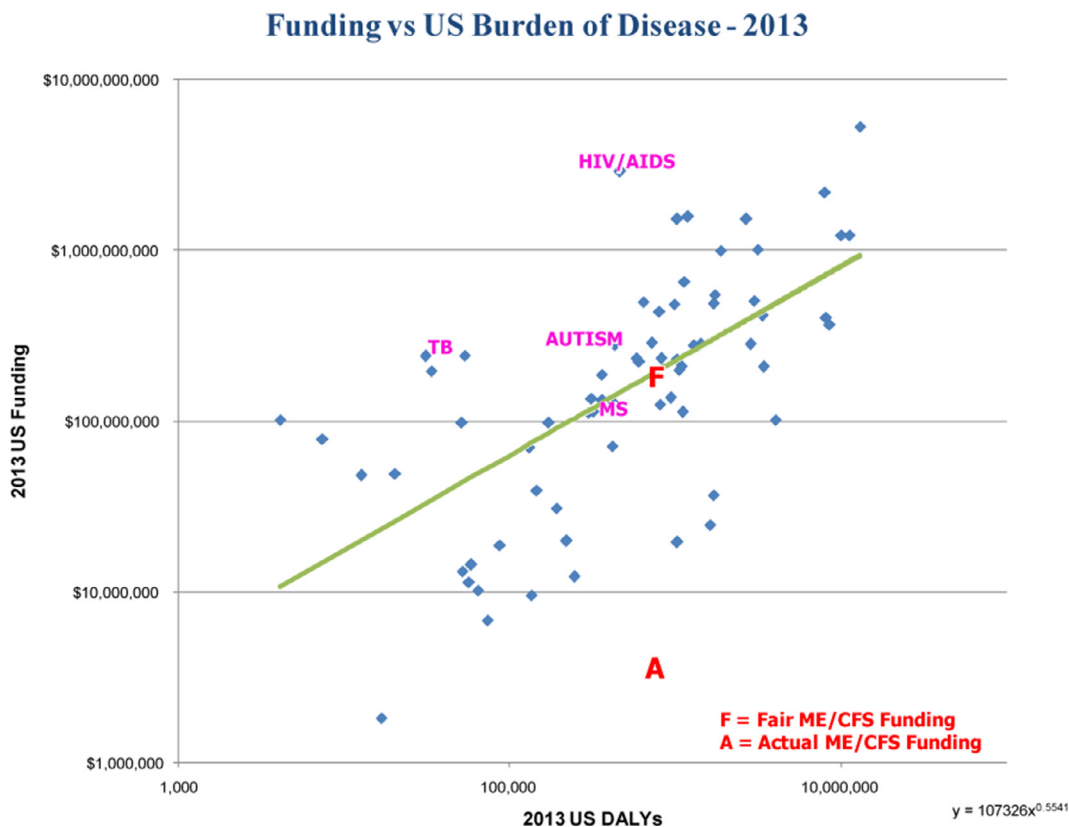
Fig. 2. NIH funding versus burden of disease (US) for major diseases. Funding of ME/CFS is uniquely low relative to the impact on quality of life (Dimmock et al., 2016).
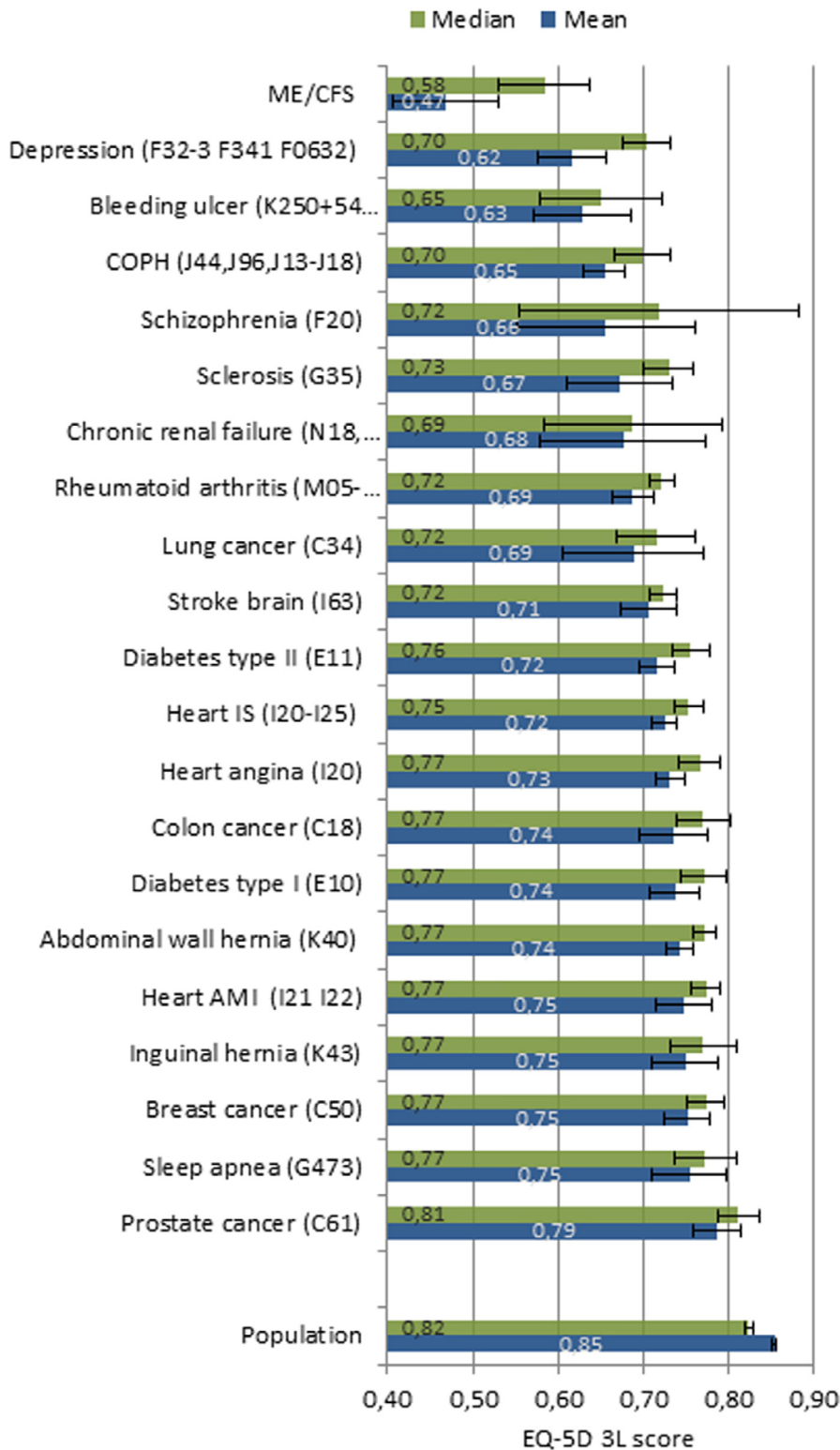
Fig. 3. Mean and median Health-Related Quality of Life (QOL) for conditions ranked by mean QOL. Score is increasing with QOL. The left (right) of the graph represents lower (higher) QOL scores. ME/CFS patients had mean QOL of 0.47 after controlling for gender, age, and co-morbidity. Controlling for these factors changed the level, but not the order of the conditions (Hvidberg et al., 2015).

more judgmental language toward some diseases. That both the LASSO-logistic regression and the dictionary analysis were comparable attests to the robustness of the results.

Our analysis shows 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 rigorous evidence for disease severity and an assumption that ME/CFS patients do not differ substantially in type compared to patients with similar conditions. Attitudes toward ME/CFS patients are comparatively more negative than those toward 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.

On forums discussing patient care, doctors discuss common treatments or strategies, as well as frustrations related to managing particularly complex diseases, but these discussions often include negative judgements about patients as well. Prejudices toward patients can harm the doctor-patient relationship by damaging patient trust and through lowering quality of care. Croskerry (2003) has discussed the way cognitive biases in physician judgements lead to diagnostic errors. In the economics literature, incorrect beliefs are not consistent with rational expectations and are a market inefficiency. If expectations were rational, assumptions about the distribution of agent type throughout a population would be correct on average. Incorrect beliefs are a distortion (Sarsons, 2017; Bohren et al., 2019). In a health context, bias against a patient group may result in lower physician effort, and patients that received the lower quality of care may have greater chances of remaining ill or experiencing worse health outcomes. For example, Heyhoe et al. (2016) found negative emotions of healthcare professionals, such as anxiety, anger, and disgust, reduce patient safety. These emotions worsen cognitive performance and increase medication errors (Heyhoe et al., 2016; Iedema et al., 2009).

Forhan and Salas (2013) presented a literature review of negative attitudes and beliefs attributed to obese people and how these could prevent development of adequate treatments for obesity. Balsa and McGuire (2001) review literature on racial discrimination and belief in healthcare in the United States (US). However, studies on the physician perspective of patients are usually qualitative or conducted in laboratory settings with small surveys. We contribute to the literature by looking at the physician's attitudes, using automated methods, developing a dictionary to measure stigma in a consistent way, and considering diagnosis as a dimension of bias. Further, we use natural language processing to determine whether this occurs systemically across the data.

The economics research around implicit bias is well developed, quantified, and has been repeatedly empirically tested and demonstrated for characteristics such as gender and race (Bertrand and Mullainathan, 2004; Bertrand et al., 2005; Bertrand and Duflo, 2016), but these models are not generally applied to the category of disease. The traditional economics model of bias focuses on the firm. The hiring manager judges the type of agent and the outcome is the wage (Sarsons, 2017). We extend this economic model to a health context where the physician is the decision maker, judging the illness severity of a patient, and the outcome is the treatment decision. In this paper we consider the perspective of physicians to contribute to understanding decision-making in healthcare markets. Measuring whether beliefs are correct is not possible here at the individual level, but can be suggested at the aggregate level, which we do in this paper. We recommend further research use this framework to measure treatment outcomes relative to beliefs to detect discrimination.

Our paper contributes to behavioral economics in that it underscores the importance of beliefs. The beliefs of doctors, and the emotions that result, influence how they behave (Heyhoe et al., 2016; Iedema et al., 2009). Sarsons (2017) showed that gender influences perceptions of worker ability and this contributes to inequality in the labor market. Within a medical context, physicians' biases can similarly impact both the health and labor markets. This reduces human welfare both through decreased quality of life and lower life expectancy.

Our work adds to a growing body of literature using social media data in economics research (Allcott and Gentzkow, 2017; Antweiler and Frank, 2004; Cox et al., 2021; Jiao et al., 2020). Bollen et al. (2011) measured mood in data of Twitter posts. Bohren et al. (2018) used data from the mathematics forum Stack Exchange to measure gender discrimination. Antoci et al. (2019) found civility in Facebook discussions fostered the trust essential for economic development. These papers establish precedent for using social media data to answer economic questions. We build on this work by researching provider behavior to contribute to understanding healthcare markets.

Sampling bias is a challenge for economics research using social media data but is not prohibitive. As previously described, some of the sites used for economic research using social media data include Twitter, Stack Exchange, Facebook, and the Economic Job Market Rumors forum (Bollen et al., 2011; Bohren et al., 2018; Antoci et al., 2019; Wu, 2017). Each of these four sites has unique population biases. Despite selectivity concerns, these papers provide valuable economic insight.

The remainder of the paper is organized as follows. Section 2 and 3 describes the literature and the data, Section 4 presents the results, while Section 5 concludes on the findings of the paper.

2. Data

Reddit is a global online forum with more than 430 million monthly active users. The site is split into communities known as subreddits, each of which covers a different topic (Widman and Nicol, 2018). Users are primarily from the United States (US), United Kingdom (UK), and Canada. The US makes up 53.9% of visitors, while the UK and Canada make up 8.2% and 6.3% of visitors. With more than 50 million daily users, Reddit is one of the most visited websites globally (Similarweb, 2022).

For this study, we use text of comments from the "Medicine" subreddit (r/medicine), a forum for healthcare professionals. The welcome page[4] describes the forum as a place for "physicians and other medical professionals from around the world to talk about the latest advances, controversies, ask questions of each other, have a laugh, or share a difficult moment." This is one of the largest groups of physicians commenting anonymously online. The forum, founded in 2008, has more than 358,000 members. In 2020, there were 1,018,776 practicing physicians in the US (Young et al., 2021).

We obtained nearly 50,000 observations from the site API. User information does not include their gender or location. The forum is designed by and for health care professionals as specified in the main text. The forum has strict rules enforced by a moderation team. The rules users must follow to participate in discussion include restrictions like the following:

- All submissions must be made by users who have set "flair," indicating 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.
- No personal agendas or "single topic" commentary.
- Users must protect patient confidentiality.[5]

The rules explain that 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 posts mostly in diabetes subreddits, then they may not post about diabetes on r/medicine. If a user posts on the forum who has comments or posts in their history that contradict the posts on the forum, especially regarding medical background, this is against the forum rules. These rules are used to ban posts, comments, or users and prevent them from continuing to participate.

Furthermore, we removed comments made by users who were not medical professionals using the account tags. Posts appear to include more medical students and residents. This is beneficial as understanding the attitudes of young doctors is important for recruiting practitioners who specialize in certain types of diseases. Often in social media data the background of the user is not available; the ability to sort by occupation and the investment in moderation of the forum makes this data unique.

Users post anonymously, meaning text should capture doctors' actual beliefs to a greater extent than using their public comments, notes in medical records, or behavior in lab settings. It is possible to determine, within this group, whether they talk about different conditions with different language. Unlike comments made in formal settings, social media data is rawer and richer. It is possible to get more information and it is more likely to pick up signals compared to the contexts of interviews or lab environments where doctors may self-censor. As Reddit is one of the most visited sites in the world, the data from provider conversations is invaluable for analysis of medical culture (Widman and Nicol, 2018; Similarweb, 2022).

There is not extensive research using data from Reddit. The studies that analyzed Reddit data, focused on different forums such as disease-specific subforums (Park et al., 2018) or examined Reddit users as a general group (Park and Conway, 2018; Record et al., 2018). These studies examined the attitudes of patients or the public. We did not find studies using Reddit to analyze naturally occurring language of doctors toward patients. In our study, stigma refers to negative beliefs held by physicians identified using text data downloaded using the site API.

We use ME/CFS and compare it with 20 other major diseases selected from the study, *The Health-Related Quality of Life for Patients with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS)*, which compared ME/CFS to other medical conditions, and ranked them by quality of life. The nature of the disease has hidden it from the public and medical professionals, potentially contributing to misconceptions. Furthermore, research by Geraghty and Esmail (2016) has shown that ME/CFS patients face high rates of misdiagnosis, both of misdiagnosing ME/CFS inappropriately in patients with another chronic disease and of not diagnosing patients when necessary. Skepticism and uncertainty may contribute to this dynamic. Given this, we hypothesize that ME/CFS patients face a high degree of bias that is detectable in language.

The goal was to measure whether any apparent stigma was representative of incorrect beliefs, which is measured relative to patient severity. The Center for Disease Control (CDC) describes ME/CFS as a long-term "disabling and complex illness" that "affects many body systems" (CDC, 2021). This disease impacts 2.5 million Americans and is more common than multiple sclerosis (MS), Parkinson's disease, or AIDS (IOM, 2015). If ME/CFS patients face more negative attitudes from medical professionals, despite being more severely ill than other conditions in the sample this gives evidence for incorrect beliefs. For this, we use the measures of quality of life. Fig. 3 shows the results and diseases list. When ranked in terms of quality of life (QOL), ME/CFS patients have lower average QOL than that of many other major diseases. ME/CFS is more disabling than stroke, diabetes, renal failure, heart failure, or lung cancer (Hvidberg et al., 2015). Geraghty and Blease (2018) found that, despite the low QOL scores of ME/CFS patients, many medical professionals do not view ME/CFS as a severe medical condition and are not aware of the biomedical literature. Bowen et al. (2005) surveyed attitudes of English general practitioners. The survey found 48% of providers lacked confidence to diagnose ME/CFS patients and 41% lacked confidence to treat.

To address the 21 diseases in the study, ME/CFS and 20 other major diseases, we pulled data from the Reddit API searching by disease name and common synonyms. This returned all submissions available in the database that had enough content in the title, post text, or comments, related to the disease name the API returned as relevant to that disease. Our sample has 11,128 observations including 426 posts and 10,702 comments by physicians.

---

[4] https://www.reddit.com/r/medicine/.

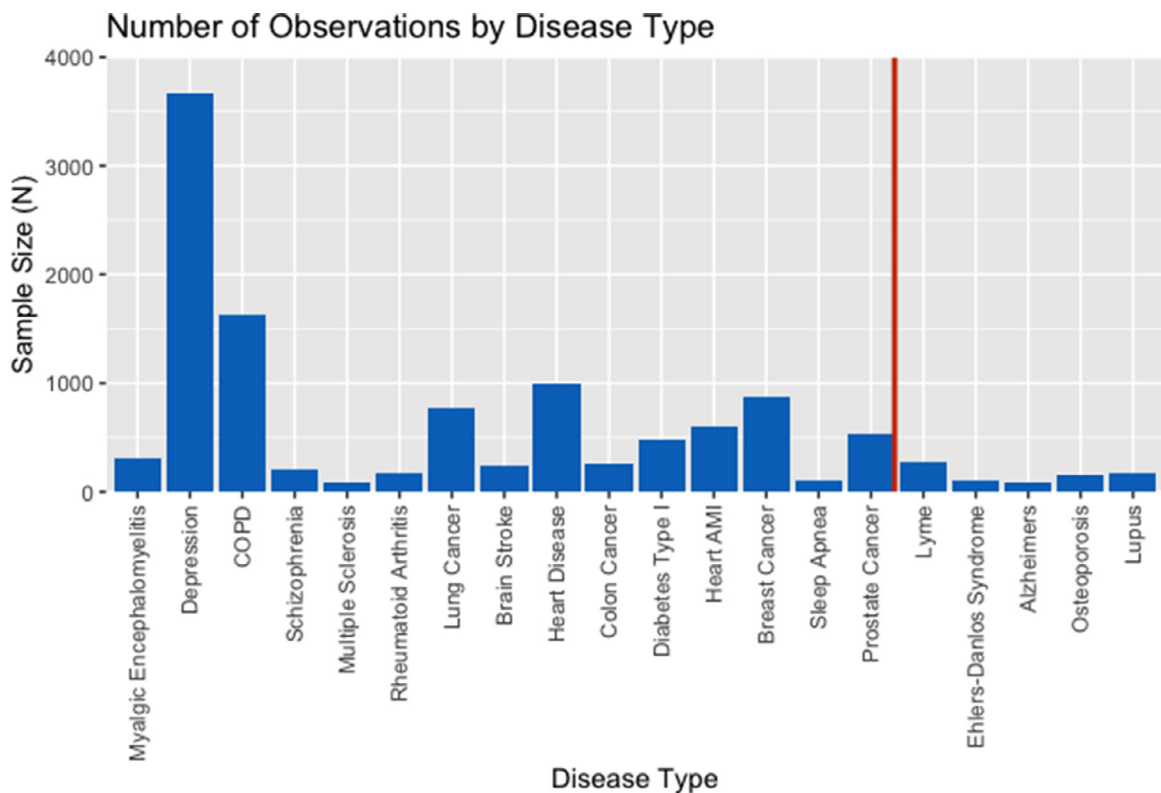[5] *r/medicine*, https://www.reddit.com/r/medicine.

Fig. 4. 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 5.3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Investigating this research question required collecting novel data. We wrote a script to access the forum database and cleaned the returned data for the purpose of this study.

The sample initially included 26 diseases listed in Fig. 3 on page page 27. We dropped five of these diseases 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. These diseases are not interesting enough for any reason to be worthy of discussion. They are likely to be uncontroversial and have relatively low stigma. An exception being Diabetes Type II which did not have enough data because it did not have enough differentiating characteristics to separate it from Diabetes Type I.

Submissions that were returned both for a disease name and its synonym were deleted. Data was pulled for synonyms of disease names as well, to gather posts that may appear under another name. Next, we deleted duplicate entries as well as submissions returned for multiple diseases to train the model on unique submissions relevant to only one disease. As a robustness check, we collected data for an additional five diseases. These additional diseases are those to the right of the red line in Fig. 4 on page page 28.

Once connected to the Reddit API, we wrote a script to collect all the posts related to each disease, all the posts' comments and all the sub-comments. A single post could have hundreds of comments branching off from the initial comments that each broke into their own comment trees. The largest number of comments for any post in this data set was 800. Fig. 4 shows the total number of observations for each condition.

Due to the large number of words, we cleaned and processed the text using steps aimed at reducing dataset dimensionality. In Appendix A we present the dimensionality reduction algorithms used to process the text to prepare it for analysis.

Below are qualitative examples of the effect that the paper attempts to measure. The following comments were left by doctors discussing ME/CFS. It appears medical professionals from the US wrote many of the comments. 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 US.[6]

---

[6] 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 US and the name used by doctors and researchers working there (Rowe et al., 2017).

The following are examples of comments:

> Chronic Lyme, adrenal Fatigue, CFS, Ehlers-Danlos[7], Fibromyalgia[8] , SIBO[9] and if you really get a believer, Morgellons.[10] 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

Further comments are illustrated in Appendix C.

3. Method of analysis

After we collected the data, we transformed the variables into a numerical document-term matrix, we standardized them and used LASSO-logistic regression to determine which words had the greatest predictive power for disease category.

In high-dimensional cases, with an X matrix that contains thousands of variables, it is not clear which variables are relevant. LASSO can help by discarding variables that are not useful. In text data, there is often a high degree of multi-collinearity among the words. The LASSO penalty is indifferent to the choice among a subset of predictors that are correlated (Hastie et al., 2009). Redundant attributes that do not add additional information are removed by setting their coefficients to zero. This results in a model with fewer features that is easier to interpret (Fonti, 2017).

The estimator for the LASSO-logistic regression is:

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

and:

$$\|\beta\|_1 = \sum_{j=1}^{p} |\beta^j| \tag{2}$$

The loss function ρ from the equation above is:

$$\rho(\beta) = -y(\sum_{j=0}^{p} \beta_j x^j) + log(1 + exp(\sum_{j=0}^{p} \beta_j x^j)) \tag{3}$$

Since the absolute value is penalized, we standardized the variables. The size of the penalty parameter in the constraint, λ, was chosen through 5-fold cross-validation. The resulting $\hat{\beta}$ for each attribute is biased on average but has smaller variance leading to more reliable predictions (Wu, 2017). For more details about LASSO see Hastie et al. (2009).

We used a binomial response variable to detect the words with greatest predictive power for ME/CFS related posts. This is a classification problem, which is a subset of machine learning problems with a discrete (or categorical) response variable (Hansen, 2018). Here, the category is disease type. The binary model is used to determine whether a given word is predictive of ME/CFS (1) or not (0). The outcomes indicate whether the post was returned by the Reddit API when searching for the disease name as well as common synonyms. Posts marked with the outcome ME were those the API returned for chronic fatigue syndrome, myalgic encephalomyelitis, CFS, and ME/CFS.

This application of LASSO returned a list of the most relevant words for ME/CFS and 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 sign of the coefficients show the change in log-odds[11] of a comment being about ME/CFS for a unit change in the standard deviation of the word frequency.

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:

$$\pi_i = P(ME_i = 1|x_i) = \frac{exp\{\beta_0 + x_i^T \beta\}}{1 + exp\{\beta_0 + x_i^T \beta\}} \tag{4}$$

---

[7] Ehlers-Danlos syndrome (EDS) is a group of conditions that affect connective tissue.

[8] Fibromyalgia syndrome (FMS), is a condition that causes long-term, widespread pain in the body among many other symptoms.

[9] Small Intestinal Bacterial Overgrowth.

[10] *MedicineNet* 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" (Dryden-Edwards, 2020).

[11] logarithm of the odds ($\frac{\pi}{1-\pi}$).

$$1 - \pi_i = P(ME_i = 0|x_i) = \frac{1}{1 + exp\{\beta_0 + x_i^T \beta\}} \tag{5}$$

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

$$P(ME_i|x_i) = P(ME_i = 1|x_i)^{ME_i} \times P(ME_i = 0|x_i)^{1-ME_i} \tag{6}$$

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

$$l_n(\beta) = log(\Pi_{i=1}^n P(ME_i|x_i)) = \sum_{i=1}^n ME_i(\beta_0 + x_i^T \beta) - log(1 + exp(\beta_0 + x_i^T \beta)) \tag{7}$$

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 $p \times 1$ vector of word counts that are stacked into the X-matrix which is $n \times p$ doc-term-matrix.

Words that are not selected by the model have coefficients of zero. These words are not relevant 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. Here, k denotes a specific word from the vocabulary of size p. For the remaining words, the marginal effect of the kth word is the difference between the probability of having one extra occurrence of that word and having the actual number of occurrences:

$$P(ME_i = 1|x_{i,(-k)}, x_{i,k} + 1) - P(ME_i|x_{i,(-k)}, x_{i,k}))$$
$$= \frac{1}{n} \sum_{i=1}^n \{P(ME_i = 1|x_i) \times P(ME_i = 0|x_i)\}\hat{\beta}_k \tag{8}$$

The coefficients express the marginal effect of each word selected for inclusion in the model, representing the change in log-odds that the comment is related to ME/CFS if the frequency of the word increases by one standard deviation (Leeper, 2018).

To further the analysis, we ran binomial regressions for two additional disease categories: depression and autoimmune disease. Depression was chosen to compare the results for ME/CFS to those of another stigmatized condition and to reveal differences in language between ME/CFS and mental illness. Similar to ME/CFS, depression patients also have extremely low average quality of life scores. Comparing the two diseases will reveal if discussions regarding ME/CFS have more aggressive language beyond that directed toward mental illness patients. The autoimmune category includes both multiple sclerosis and rheumatoid arthritis as autoimmune diseases have characteristics in common with ME/CFS.[12] Patients from similar demographics with similar symptoms such as MS and RA compared to ME/CFS should have similar characteristics. Thus, there is no reason to expect that ME/CFS patients are less agreeable than MS or RA patients. If ME/CFS patients face more negative attitudes, this provides evidence of bias on the part of medical providers.

Using the two additional binomial regressions, we check whether ME/CFS is discussed with tone that conveys less sympathy or urgency than similar conditions. MS and RA are classified as physical illnesses which makes for helpful comparison with depression.

In Section 5, we measured the tone of the text using the General Inquirer. This is a word list developed for sentiment analysis where each word is tagged to an emotion. General Inquirer classifies over 8000 words into 182 emotional categories. As stigma was not among the emotion categories, we created a custom dictionary to measure this. The use of this stigma dictionary is discussed further in Section 5.

We chose LASSO regression because, in the event multiple variables are collinear, LASSO will select one variable from the group and ignore the others, simplifying the model. As a further robustness check we have included results for the elastic-net regression, which combines qualities of both LASSO and ridge regressions. We used the standard alpha of 0.5. The elastic-net results and equations are in Appendix B.

4. Results

The results of the binomial case are shown in Table 1, which lists 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, discussed further in Section 5 are in bold and words that are medical terminology are in italics. Commonly understood terms were not classified as medical, but only technical medical terms and jargon. The appearance of medical terminology is used as a proxy

---

[12] In Jason et al. (2011) and Jason et al. (2017) the authors point out that MS and ME/CFS have similar demographics, and mostly impact women. Additionally, they have a severe impact on quality of life and exhibit similar symptoms. Jason et al. (2011) found that when using symptoms alone, it is difficult to distinguish ME/CFS from MS for epidemiological research. They are also potentially part of the same disease category. There is growing support for the theory that ME/CFS is an autoimmune disease (Nakatomi et al., 2014; Montoya et al., 2017; Davis, 2017). As MS alone did not have a large enough sample of comments, it was combined with another autoimmune disease from the original sample, rheumatoid arthritis, which represents a similar demographic and disproportionately impacts women.

Table 1
LASSO: Words with strongest predictive power for **ME/CFS**.

Words	Marginal Effect	Words	Marginal Effect	Words	Marginal Effect
afflict	69.41	*microbiome*	4.44	debilitate	1.59
skeptic	66.90	**disprove**	4.38	rule	1.55
jamison	55.94	rename	4.37	subreddit	1.53
jake	49.63	*cidp*[f]	4.18	aerobic	1.42
unproductive	39.89	**psychosomatic**	4.17	**dismiss**	1.42
sibo[a]	38.81	**nothingness**	3.43	**observations**	1.41
cyclophosphamide[b]	27.93	reproducible	3.24	cherrypicked	1.41
ehlersdanlos	20.49	methodology	3.21	illnesses	1.40
morgellons	15.41	overgrowth	3.16	rmedicine	1.39
towel	14.53	jamisons	3.07	nerve	1.35
mudphud[c]	12.75	relay	2.90	**degrade**	1.32
candida	11.69	concise	2.89	**remove**	1.31
neuropsychiatry	10.81	*ferritin*	2.80	science	1.31
nonpsychiatric	10.35	scientist	2.39	researchers	1.28
instagram	9.76	*pots*[g]	2.19	**vague**	1.25
sds[d]	9.01	cis	2.16	underpowered	1.20
saga	8.57	donate	2.15	*pcps*[i]	1.20
epidemiologists	8.43	**worldview**	2.12	**agenda**	1.18
scoff	7.83	disable	2.12	sidebar	1.17
hypermobile	7.65	*savella.*	2.07	post	1.15
vtwl[e]	7.07	psychological	2.06	paperwork	1.15
somatized	7.06	*picc*[h]	1.98	link	1.09
sd	6.83	unsound	1.87	*iom*[j]	1.07
elude	6.76	**accuse**	1.79	diverse	1.06
scientists	6.30	norway	1.79	**unprofessional**	1.05
intolerance	6.24	**conspiracy**	1.78	**laziness**	1.04
exertion	5.91	scientifically	1.73	stanford	1.04
pulmonologist	5.85	favourite	1.72	illness	1.00
madeup	5.80	catchall	1.64		
documentary	5.42	ama	1.62		

Table lists features from LASSO-logistic model with greatest predictive power for ME/CFS and the marginal effect. Regression was binomial with a single outcome variable indicating 1 for ME/CFS and 0 for all other diseases. Words in **bold** are part of the stigma dictionary. Words in *italics* are medical terminology.

[a] small intestinal bacterial overgrowth
[b] cancer treatment
[c] MD PhD
[d] standard deviations
[e] ventilatory threshold
[f] chronic inflammatory demyelinating polyneuropathy
[g] postural orthostatic tachycardia syndrome
[h] peripherally inserted central catheter
[i] primary care physician
[j] Institute of Medicine

for the inverse of stigma and is useful because it is not subjective or context dependent. Discussions focused on treatment may have widespread use of medical abbreviations and jargon. This could indicate physicians take these conditions more seriously.

Table 1 shows that for ME/CFS, words with the strongest predictive power include "skeptic," "unproductive," "scoff," "morgellons," and "laziness."[13] Finding exp^{β_k} for each coefficient gives the listed marginal effects. A one standard deviation increase in the frequency of the word "scoff" will multiply the relative risk a comment is about ME/CFS by 7.83. Phrased another way, this is a 783% increase in the probability a comment is discussing ME/CFS if the word "scoff" appears in the text. The same increase in the word "laziness" increases the probability a doctor is discussing ME/CFS by 104%.[14] From these terms, it is clear the discussion is not primarily regarding treatment strategies or biomedical research. There is medical terminology in the predictive words list for ME/CFS, but the language is not extensively technical. The majority of the medical terms are words a layperson could understand such as "pulmonologist" and "epidemiologists."

The analysis was furthered with regressions for two additional disease categories: depression and autoimmune disease. The words with the most predictive power for these diseases are shown in Tables 2 and 3.

Comparing the results for ME/CFS in Table 1 to those for depression in Table 2, depression has fewer stigma words. While ME/CFS has 21 stigmatizing words, depression has only five. The size of the coefficients for the ME/CFS words are

---

[13] MedicineNet 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" (Dryden-Edwards, 2020).

[14] For comparison, in Wu (2017), the words with the greatest predictive power for whether a comment referred to a female economist included "hotter" and "attractive," while those for male economists included "philosopher" and "motivated."

Table 2

LASSO: Words with strongest predictive power for **Depression**.

Words	Marginal Effect	Words	Marginal Effect	Words	Marginal Effect
rmedicalschool	6.76	*naltrexone*	3.18	vectors	2.63
ketamine	6.09	characterization	3.16	aid-in-dying	2.62
cookies	5.62	tombstones	3.15	bureaucrats	2.62
chemicals	5.24	neymar	3.13	quest	2.61
zoloft	5.21	**unscientific**	3.05	artists	2.61
reapproved	4.59	lbgtq	3.00	avril	2.58
plumbers	4.44	movies	3.00	*ibs*	2.56
corrigans[a]	4.44	toxins	2.99	*ld*[j]	2.56
physicals	4.37	psych	2.99	himher	2.54
acgme[b]	4.29	lifes	2.98	maois	2.52
thrive	4.24	marathon	2.96	rewrite	2.52
porcelain	4.24	*implantation*	2.94	chiefs	2.51
lonely	4.16	antidepressants	2.90	collectively	2.50
unflappable	4.05	bbc	2.90	simulate	2.48
sperm	3.87	treatmentresistant	2.90	unspoken	2.48
ssri	3.81	horrendous	2.89	solo	2.45
mdd[c]	3.67	*sertraline*	2.87	rfibromyalgia	2.43
antidepressant	3.62	depressive	2.86	*benzos*	2.41
premeds	3.57	enthusiasm	2.85	*haloperidol*	2.39
indefinite	3.48	coincidence	2.85	acne	2.38
moan	3.41	clock	2.81	unfair	2.37
url	3.40	*fibroids*	2.80	*ssris*	2.33
rsuicidewatch	3.39	suicide	2.80	maid	2.33
postgraduate	3.36	debilitation	2.78	**disproven**	2.31
psychosocial	3.32	authorization	2.76	obstacle	2.31
suboxone	3.24	custom	2.75	depress	2.27
rtms[d]	3.22	gift	2.75	drag	2.26
rln[e]	3.22	*adhd*	2.75	**iffy**	2.26
rekgs[f]	3.22	*fibro*	2.75	slot	2.24
meckels[g]	3.22	**crock**	2.73	psychic	2.23
calciphylaxis[h]	3.21	humane	2.72	secretions	2.23
physio	3.20	buff	2.69	mayosi	2.21
asshole	3.20	*od*[i]	2.68		

Table lists features from LASSO-logistic model with greatest predictive power for depression and marginal effect. Regression was binomial with a single outcome variable indicating 1 for depression and 0 for all other diseases. Words in **bold** are part of the stigma dictionary. Words in *italics* are medical terminology.

[a] Corrigan's disease
[b] Accreditation Council for Graduate Medical Education
[c] major depressive disorder
[d] repetitive transcranial magnetic stimulation
[e] recurrent laryngeal nerve
[f] reconstructed electrocardiogram (rEKG)
[g] Meckel's diverticulum
[h] calcification of the small blood vessels
[i] overdose
[j] labor and delivery

overall larger than those for depression, showing the words are more strongly associated with the condition. In contrast to ME/CFS, the most predictive words for depression were more weakly associated. It has been documented that doctors have negative prejudice against mental health disorders (Jorm et al., 1999; Caldwell and Jorm, 2001). Thus, it is perhaps surprising this data suggests depression is less stigmatized than ME/CFS. Examining the medical terminology used to discuss ME/CFS and depression, Tables 1 and 2, ME/CFS has 19 predictive words in italics indicating medical jargon, while there are 25 for depression.

Comparing ME/CFS in Table 1 to autoimmune diseases in Table 3, shows much different predictive words between the two groups. For the autoimmune category, LASSO did not select any words from the stigma dictionary as predictive. The words list for ME/CFS includes 21 sigma words with predictive power compared to zero for autoimmune disease. Thus, ME/CFS is more stigmatized in this sample. There is more medical terminology in the autoimmune list including more words that refer to body systems and disease processes. For autoimmune disease, the most frequent category of words used was medical terminology and included words such as: "retrovirus," "hyperreflexiad," and "etanercept." This is significant considering that fewer words were selected for autoimmune disease overall, but medical jargon represented a higher proportion of the selected words.

The following analysis included additional cases. However, these approaches yielded similar results: in this sample of data, doctors appear to hold negative stereotypes about their ME/CFS patients, and those stereotypes are systemic.

Table 3

LASSO: Words with strongest predictive power for **Autoimmune diseases (MS and RA)**.

Word	Marginal Effect	Word	Marginal Effect
norris[a]	49.87	tolerable	2.36
immunosuppression	45.09	*tia*[j]	2.24
reno	29.43	*tb*	2.17
ocrevus	27.96	potato	2.02
pmr[b]	17.12	psychology	1.90
pan-body	16.64	game	1.86
rituxan	10.48	jeez	1.75
selfdefined	8.56	realistic	1.63
tnfalpha[c]	8.12	*sclerosis*	1.61
viruses	7.61	*virus*	1.59
retrovirus	6.56	tolerance	1.58
ppms[d]	6.23	pain	1.58
vasculitis	6.11	*cns*	1.56
hyperreflexia[e]	5.51	*nivolumab*[k]	1.41
microbiome	5.13	asses	1.41
ncv[f]	4.81	vital	1.41
jwatch[g]	4.81	accident	1.40
brandname	4.81	*jbjs*[l]	1.35
calorie	3.94	scale	1.30
gbs[h]	3.84	*etanercept*[m]	1.27
journal	3.55	*rituximab*	1.25
formulate	3.54	hypotheses	1.21
radius	3.35	rash	1.17
pilonidal[i]	2.85	migraine	1.15
chuck	2.75	*dna*	1.13
antiretrovirals	2.45	sublingual	0.94

Table lists features from LASSO-logistic model with greatest predictive power for the autoimmune diseases MS or RA and marginal effects. Regression was binomial with a single outcome variable indicating 1 for MS or RA and 0 for all other diseases. Words in **bold** are part of the stigma dictionary. Words in *italics* are medical terminology.

[a] Chuck Norris' wife was injured by radioactive dye from an MRI
[b] inflammatory disorder causing muscle pain
[c] inflammatory cytokine
[d] Primary-progressive multiple sclerosis
[e] overactive or overresponsive reflexes
[f] nerve conduction velocity
[g] NEJM Journal Watch
[h] Guillain-Barré syndrome
[i] abnormal skin growth
[j] transient ischemic attack
[k] immunotherapy
[l] Journal of Bone and Joint Surgery
[m] TNF blocker

5. Dictionary correlations tests

Gentzkow et al. (2019) illustrates that a common method of sentiment analysis is to use standardized dictionaries to measure tone, such as positivity or negativity in the text using lists of common words. One such standardized dictionary, the General Inquirer, was developed by Philip Stone at the Harvard Laboratory of Social Relations and is one of the most extensive. This dictionary includes emotions such as Positive, Negative, Affiliation (or Supportive), and Hostile.[15] However, these emotions are not specific enough to be helpful for our research question.

Simply measuring whether discussion is negative will not capture whether there is stigma; both a negative statement about a patient's character and a negative statement lamenting the severity of a disease or the incidence of mortality would show as negative. 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 using word lists alone this hostility is not distinguishable from comments that use words like "attack" to refer to aggressively treating a disease.

The emotions previously classified in the General Inquirer would not measure bias. Thus, instead, we manually classified words into a dictionary for stigma. We conduct three tests using dictionary techniques on this custom stigma list. The first

---

[15] The General Inquirer, accessed April 23, 2019, http://www.wjh.harvard.edu/~inquirer/.

Table 4

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
Stigma	Stigma also in General Inquirer (maintained 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 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. There are two stigma dictionaries, General Stigma dictionary and the Original Stigma dictionary. This table shows: 1) Example words from the General Stigma dictionary. This only includes that appear both in the Original Stigma dictionary and the General Inquirer and 2) Example removed words from the Original Stigma dictionary. These are more specific stigma words that did not appear in the General Inquirer.

used a dictionary of words classified as stigma or not. The second used a second dictionary including words only from the stigma list that also appeared in the General Inquirer. The final test conducted a blinded test using five additional diseases on both previous dictionaries. None of the text data for these diseases had been viewed prior to the analysis. These tests are described below. These dictionary correlations show effects similar to the LASSO regressions. The two techniques reinforce each other, which verifies their robustness.

5.1. Original stigma words test

It is possible that diseases with negative stereotypes have more negative words. In this scenario, it would appear 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 have 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. We then checked the correlation between this dictionary and the word frequencies for each disease. To define the stigma dictionary, we manually classified words from the first 1000 comments as either having stigma or not. As a word list for this concept does not exist, this research required the creation of a new dictionary. Judgement is inherently a part of this. To classify the words, we used the following framework: words were added to the dictionary if they referred to either personality or mood states and also if they implied imprecision or vagueness. Words that implied doubt, disbelief, hostility, or frustration were labelled with a one for stigma while all other words in the data were set to zero.

There is precedent for classifying words into categories from Wu (2017). We follow this same approach. Wu (2017) built a new dictionary from the data by classifying words as either personal or professional. She regressed the gender of the posts on the two categories. 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 selected by LASSO for ME/CFS, the stigma dictionary provides another way to confirm that these are not words used about other conditions in the sample; in particular, conditions that have similar characteristics.

Using this dictionary, we 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, it indicates a linear relationship between the stigma dictionary and that disease.

Table 5 shows the results of all three tests in one table. The first involves the upper half of this table, column one. This shows the correlations between the diseases in the sample and the stigma words.[16] Table 5 shows the correlations between a vector of the frequency counts of every word for a given disease, and a vector of ones and zeros indicating whether or not the word was tagged with stigma. The correlations were only between the words that were in both lists: the frequency matrix from the data and the words in the stigma dictionary. For brevity, Table 5 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 is occurs across all the documents in the corpus. The other diseases are filtered out, entailing that only one disease

---

[16] There are 15 diseases because five diseases from the original 20 were dropped due to lack of data, as discussed above.

Table 5
Stigma correlations.

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

Pearson correlation coefficients between the term frequencies and the stigma words shown with p-values in parentheses. Significant values in bold.

is considered at a time. Each word is an observation while the independent variable, stigma words, is a binary variable of ones and zeros, indicating if the word is in the stigma category.[17]

The result was that this new stigma dictionary was correlated with ME/CFS and not with any of the other diseases. For the first row of Table 5, 0.05 is the correlation coefficient of the word frequencies for ME/CFS with the binary stigma variable. The p-value is smaller than 0.00, indicating that it is statistically significant.

A given post or comment is likely to contain very few stigma words. For example, if a patient is called "lazy" then that word may show up only once in the comment, with the presence of this word dampened by the presence of many other non-stigma words in the comment making the correlations insignificant. Yet this outcome did not occur. The stigma dictionary was significantly correlated with whether a document was ME/CFS and not with the other diseases in the sample. The result showed that dismissive words were not only correlated with ME/CFS but were not correlated with any other major disease checked at all.

Table 6 shows the full result of the regression summarized in Table 5. This test is represented by the ME/CFS row, row one, column one of Table 6. For ME/CFS, words in the original stigma dictionary have an average frequency of 2.4 more words than non-stigmatized words. Thus, this table shows that, for ME/CFS, stigma words appear significantly more frequently than other words.

---

[17] With a binary variable as the regressor, these standard errors are not 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 6
Regression of word frequencies on stigma words .

ME/CFS		Stigma Words	Stigma and General Inquirer Words
		2.403***	1.559***
		(0.498)	0.446
	constant	0.805***	0.808***
		(0.035)	(0.035)
	R^2	0.0023	0.0012
	Adj. R^2	0.0022	0.0011
Lyme Disease		0.912*	0.649
		(0.407)	(0.364)
	constant	0.567***	0.567***
		(0.028)	(0.028)
	R^2	0.0005	0.0003
	Adj. R^2	0.0004	0.0002
Ehlers-Danlos Syndrome (EDS)		0.206	0.165
		0.199	0.178
	constant	0.252	0.252***
		0.014	0.014
	R^2	0.0001	0.0001
	Adj. R^2	0.0000	-.0000

Footnote: Signif. codes: '***' if <0.001, '**' if <0.01, '*'if <0.05, '.' if <0.1

5.2. Generalized stigma words test

We created a generalized stigma dictionary from more common words and checked the correlations again. This second dictionary joined the previous stigma dictionary with the General Inquirer. Words that did not appear in both were removed.

The purpose of using only stigma words that do appear in the General Inquirer is to have a methodology that reduces the potential for bias. As these words were chosen and classified by an independent team, the use of this dictionary should improve objectivity of the results. This goal was also furthered by the exclusion of rare words.

Table 4 lists example General Inquire words, maintained words, and removed words. Stigma words that do not appear in the General Inquirer are less common words and are likely specific to a particular context. For this second test, only the more common words were used. When the correlations were run with the stigma words that also appear in the General Inquirer, the result is similar to the previous test. These results are in the second column of Table 5, represented by the first 15 rows. Similar to the first test, this second, more general stigma dictionary was again correlated with ME/CFS but not any of the other diseases. The correlation coefficient for ME/CFS was 0.04; the coefficient remained significant, though slightly lower in size than in the previous test. This is visible when comparing the columns in the first row of Table 5. This reflects that the new dictionary contains more general words.

Table 6 shows the detailed results from the regression. The second dictionary with only words in both the stigma and General Inquirer dictionary was associated with 1.6 additional words than words that were not in the dictionary. Previously, the coefficient was higher, at 2.4. Now, this coefficient is lower because context-specific words have been filtered out and only generic words remain but still remain significant. For ME/CFS, these stigmatizing words appeared in discussion more frequently. As before, this was not the case for the other major diseases.

5.3. Test with five additional diseases

Do these results hold for other conditions? It is possible the previous results appeared due to chance or bias in the dictionary creation process. To address this, we repeated the previous steps on five more diseases for which none of the data had been viewed as a robustness check on the stigma correlation results. If our hypothesis holds, then it is clear the results are not due to random chance.

The additional diseases are hypothesized to bring about similar levels of stress and similar manifestation of emotional symptoms in patients to compare for robustness. The five new diseases were: Lyme disease, Ehlers-Danlos syndrome (EDS), Alzheimer's disease, osteoporosis, lupus.[18] Next, a test of correlation checked for stigma in the tone without viewing any of the text data for those conditions. This test was completed before the multinomial regression in the previous section and was truly blinded.

The results are shown in Table 5 and details given in Table 6. The results matched our expectations in four out of five of the additional diseases. The original stigma dictionary was correlated with one of the two diseases that were expected

---

[18] Lyme disease and EDS were chosen because they were hypothesized to face dismissive attitudes from healthcare providers. Alzheimer's disease and Osteoporosis were chosen because they are diseases that affect older patients who may experience age-related stigmas. We included these diseases to determine whether the hypothesized conditions (ME/CFS, Lyme, and EDS) had stigma in their tone that followed a different pattern from potential ageism.

to be stigmatized (Lyme disease, but not EDS) and none of the 3 diseases expected not to be stigmatized (Alzheimer's disease, osteoporosis, and lupus). Of note, this dictionary, represented by the first column of Table 5 was correlated with both ME/CFS and Lyme and not with the other diseases. Except for EDS, our expectations matched the outcome in all cases. The coefficients are low because stigma words are relatively rare words compared to other types of words including other common emotions.

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 decreased from 5% to the 10% significance level using the more generic stigma words. The results for EDS were, again, contrary to expectations in that the word frequencies were not significantly correlated with either of the two stigma dictionaries.

A full OLS regressions are reported in the first column of Table 6. For purposes of comparison, between the diseases that were significantly related to stigma words and those that were not, Ehlers-Danlos syndrome (EDS) was added to the table. For Lyme disease, the word frequencies were lower than those found for ME/CFS - 0.9 and 0.6 for the stigma and stigma/General Inquirer dictionary - but were significant. For EDS, these effects were smaller and did not have statistical significance just as before.

6. Discussion and conclusions

This paper used machine learning and natural language processing tools to examine language used by physicians on medical forums and determine whether their language indicates existence of stigma against certain medical conditions. To answer this question, we create a new dataset from an online forum, Reddit using the sub-direct of the forum medicine r/medicine; to the best of our knowledge, this site has not previously been used to measure beliefs of healthcare professionals towards patients. Using this novel data collected from the online discussion forum, we performed the LASSO-logistic regression to determine the words that were most predictive of conversations related to specific diseases. We controlled for disease type, prevalence, severity, and gender by comparing diseases that shared similar characteristics and discovered that attitudes were more negative towards some conditions. We also created a dictionary to detect biased attitudes and found it was correlated with two of three diseases hypothesized to have stigma, while not correlated with any of diseases that were hypothesized not to have stigma. This provided additional evidence for that revealed with the LASSO analysis.

There are limitations to our approach. Given that this is an anonymous online forum, specific demographic characteristics such as age, gender, and location are not available. Despite attempts by platform designers to police accounts, it is not possible to verify self-disclosed details. This is a challenge of social media research (Ruths and Pfeffer, 2014). Anonymity is beneficial because it presents an opportunity to observe users' true opinions. However, we need to keep in mind those writing are a proxy of the overall physician population. Additionally, we cannot show discrimination, without measurement of individual characteristics of the patients and a way of measuring whether the biases maintained about the individual patients were correct. This is possible in a lab setting, which would lose the benefit of anonymous information, but would allow for more control over experimental design.

Despite these limitations, we were able to find evidence for negative beliefs toward some medical conditions and that beliefs were still disproportionately negative when comparing conditions of similar severity and prevalence. ME/CFS, depression, and Lyme disease were discussed with more negative language than the other diseases in the set. The results for ME/CFS included over four times more negative words than the results for depression.

If agents were perfectly rational, medical decisions would be objective, and healthcare resources would be distributed optimally to areas with the greatest need. Sarsons (2017) demonstrates that incorrect bias is a market inefficiency because it is not consistent with rational expectations. If expectations were rational, assumptions would not be distorted by incorrect beliefs.

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 show that the word choice of medical providers and the attitudes these words represent can distinguish between different medical conditions.

The economics literature has shown that systemic biases in judgment can lead to incorrect stereotypes concerning particular groups (Bordalo et al., 2016). These incorrect beliefs have negative welfare implications and are inefficient (Bohren et al., 2019; Sarsons, 2017). This paper contributes to the existing research by studying the attitudes towards specific diseases using natural language processing techniques. If medical professionals hold inappropriately biased judgments about specific patient groups, this can lead to inefficiencies both in research and treatment. Biases are detectable in language and this has real economic implications. Language is an indicator of emotions and group beliefs. The fields of psychology and behavioral economics have shown these beliefs 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 our approach here.

Bias toward patients harms the doctor-patient relationship and limits patient trust. Furthermore, negative attitudes resulting in the delay or denial of treatment, or delaying investigation into symptoms, can prevent improvements in these

patients' health. These findings regarding different attitudes show that medical decisions regarding research funding and patient care may not be objectively determined. The attitudes of medical professionals may play a disproportionate role.

There is evidence ME/CFS patients face disproportionate negativity. We take as given that patients with comparable characteristics such as autoimmune diseases should be broadly similar; thus, ME/CFS patients would not have substantially different personality characteristics in that they are not less agreeable, and the severity is comparable as well. We do not have individual data on the patients discussed, but is it sensible to assume this holds on average. Biomedical research has shown the disease has a physical cause and quality of life surveys have shown the illness is severe. As the data support a comparative conclusion that ME/CFS patients are viewed differently, our research provides quantitative evidence that some patients face inappropriate preference-based biases or biases due to incorrect beliefs.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2022.09.023.

References

American College of Rheumatology: Prevalence Statistics (2016). Technical report, American College of Rheumatology.

NIH: Estimates of Funding for Various Research, Condition, and Disease Categories (RCDC) (2016). Technical report, NIH.

Allcott, H., Gentzkow, M., 2017. Social media and fake news in the 2016 election. Journal of Economic Perspectives 31 (2), 211–236. doi:10.1257/jep.31.2.211. http://www.aeaweb.org/articles?id=10.1257/jep.31.2.211

Antoci, A., Bonelli, L., Paglieri, F., Reggiani, T., Sabatini, F., et al., 2019. Civility and trust in social media. Journal of Economic Behavior & Organization 160, 83–99.

Antweiler, W., Frank, M.Z., 2004. Is all that talk just noise? The information content of internet stock message boards. J Finance 59 (3), 1259–1294.

Balsa, A.I., McGuire, T.G., 2001. Statistical discrimination in health care. J Health Econ 20 (6), 881–907.

Bandiera, O., Hansen, S., Prat, A., Sadun, R., 2017. CEO Behavior and Firm Performance. Technical Report. National Bureau of Economic Research doi:10.3386/w23248. Cambridge, MA

Belloni, A., Chernozhukov, V., Hansen, C., 2011. Inference for high-dimensional sparse econometric models. arXiv. ArXiv: 1201.0220

Bertrand, M., Chugh, D., Mullainathan, S., 2005. Implicit discrimination. American Economic Review 95 (2), 94–98.

Bertrand, M., Duflo, E., 2016. Field experiments on discrimination. National Bureau of Economic Research doi:10.3386/w22014.

Bertrand, M., Mullainathan, S., 2004. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. Am Econ Rev 94 (4,), 991–1013.

Bohren, A., Imas, A., Rosenberg, M., 2018. The language of discrimination: using experimental versus observational data. AEA Papers and Proceedings 108, 169–174.

Bohren, A., Imas, A., Rosenberg, M., 2019. The dynamics of discrimination: theory and evidence. Am Econ Rev 109 (10), 3395–3436.

Bollen, J., Mao, H., Zeng, X.-J., et al., 2011. Twitter mood predicts the stock market. J Comput Sci 2 (1), 1–8.

Bordalo, P., Coffman, K., Gennaioli, N., Shleifer, A., 2016. Stereotypes. Q J Econ 131 (4), 1753–1794.

Bowen, J., Pheby, D., Charlett, A., McNulty, C., et al., 2005. Chronic fatigue syndrome: a survey of GPs' attitudes and knowledge. Fam Pract 22 (4), 389–393. doi:10.1093/fampra/cmi019.

Caldwell, T.M., Jorm, A.F., 2001. Mental health nurses' beliefs about likely outcomes for people with schizophrenia or depression: a comparison with the public and other healthcare professionals. Aust N Z J Ment Health Nurs 10 (1), 42–54.

CDC (2021). What is ME/CFS?https://www.cdc.gov/me-cfs/about/index.html.

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.

Chinco, A., Clark-Joseph, A.D., Ye, M., 2018. Sparse signals in the cross-section of returns. J Finance 74 (1), 449–492.

Cox, G.W., Fiva, J.H., Smith, D.M., Sørensen, R.J., 2021. Moral hazard in electoral teams: list rank and campaign effort. J Public Econ 200, 104457.

Croskerry, P., 2003. The importance of cognitive errors in diagnosis and strategies to minimize them. Academic Medicine: Journal of the Association of American Medical Colleges 78 (8), 775–780.

Davis, M. (2017). Is ME/CFS an autoimmune disease?https://www.youtube.com/watch?v=SbnCMOHOL28.

Dimmock, M., Mirin, A., 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).

Dryden-Edwards, R. (2020). Morgellons Disease Treatment, Symptoms & Causes. https://perma.cc/93CB-5B5R.

Fonti, V., 2017. Feature selection using LASSO. Research Paper in Business Analytics, VU Amsterdam.

Forhan, M., Salas, X.R., 2013. Inequities in healthcare: a review of bias and discrimination in obesity treatment. Can J Diabetes 37 (3), 205–209.

Gentzkow, M., Kelly, B., Taddy, M., 2019. Text as data. J Econ Lit 57 (3), 535–574.

Geraghty, K., Blease, C., 2018. Myalgic encephalomyelitis/chronic fatigue syndrome and the biopsychosocial model: a review of patient harm and distress in the medical encounter. Disabil Rehabil.

Geraghty, K., Esmail, A., 2016. Chronic fatigue syndrome: is the biopsychosocial model responsible for patient dissatisfaction and harm? British Journal of General Practice 66 (649), 437–438.

Goddu, A.P., O'Conor, K.J., Lanzkron, S., Saheed, M.O., Saha, S., Peek, M.E., Haywood Jr, C., Beach, M.C., et al., 2018. Do Words Matter? Stigmatizing Language and the Transmission of Bias in the Medical Record. Journal of General Internal Medicine 33 (5), 685–691. doi:10.1007/s11606-017-4289-2. http://www.springerlink.com/content/120414/

Hansen, S. (2018). Machine Learning Methods for Economists: Introduction.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Springer.

Heyhoe, J., Birks, Y., Harrison, R., O'Hara, J.K., Cracknell, A., Lawton, R., 2016. The role of emotion in patient safety: are we brave enough to scratch beneath the surface? J R Soc Med 109 (2), 52–58.

Hogan, B., 2018. Social media giveth, social media taketh away: facebook, friendships, and APIs. International Journal of Communication 12, 592–611.

Hvidberg, M.F., Brinth, L.S., Olesen, A., Petersen, K., Ehlers, L., 2015. The health-related quality of life for patients with myalgic encephalomyelitis / chronic fatigue syndrome (ME/CFS). PLoS ONE 10 (7). e0132421–e0132421

Iedema, R., Jorm, C., Lum, M., et al., 2009. Affect is central to patient safety: the horror stories of young anaesthetists. Social Science & Medicine (1982) 69 (12), 1750–1756.

IOM, 2015. Report Brief: Redefining an Illness. Technical Report. Institute of Medicine of the National Academies.

Jason, L.A., Ohanian, D., Brown, A., Sunnquist, M., McManimen, S., Klebek, L., Fox, P., 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., Plioplys, S., 2011. A screening instrument for chronic fatigue syndrome. J. Chronic Fatigue Syndr. 3 (1), 39–59.

Jiao, P., Veiga, A., Walther, A., et al., 2020. Social media, news media and the stock market. Journal of Economic Behavior & Organization 176, 63–90.

Jorm, A.F., Korten, A.E., Jacomb, P.A., Christensen, H., Henderson, S., 1999. Attitudes towards people with a mental disorder: a survey of the Australian public and health professionals. Aust N Z J Psychiatry 33 (1), 77–83.

Leeper, T. J. (2018). Interpreting Regression Results using Average Marginal Effects with R's margins,. (p. 32).

Lucca, D., Trebbi, F., 2009. Measuring central bank communication: an automated approach with application to FOMC statements. National Bureau of Economic Research doi:10.3386/w15367.

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., 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.

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., In-aba, M., Kuratsune, H., Watanabe, Y., 2014. Neuroinflammation in patients with chronic fatigue syndrome/myalgic encephalomyelitis: an 11c-(r)-PK11195 PET study. J. Nucl. Med. 55 (6), 945–950.

Park, A., Conway, M., 2018. Tracking health related discussions on reddit for public health applications. AMIA Annual Symposium Proceedings 2017, 1362–1371.

Park, A., Conway, M., Chen, A.T., et al., 2018. Examining thematic similarity, difference, and membership in three online mental health communities from reddit: a text mining and visualization approach. Comput Human Behav 78, 98–112.

Pilkington, K., Ridge, D.T., Igwesi-Chidobe, C.N., Chew-Graham, C.A., Little, P., Babatunde, O., Corp, N., McDermott, C., Cheshire, A., et al., 2020. A relational analysis of an invisible illness: ameta-ethnography of people with chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) and their support needs. Social Science & Medicine 265. doi:10.1016/j.socscimed.2020.113369. 113369–113369

Record, R.A., Silberman, W.R., Santiago, J.E., Ham, T., 2018. I sought it, i reddit: examining health information engagement behaviors among reddit users. J Health Commun 23 (5), 470–476.

Rowe, P.C., Underhill, R.A., Friedman, K.J., Gurwitt, A., Medow, M.S., Schwartz, M.S., Speight, N., Stewart, J.M., Vallings, R., Rowe, K.S., 2017. Myalgic en-cephalomyelitis/chronic fatigue syndrome diagnosis and management in young people: a primer. Front Pediatr 5. 121–121

Ruths, D., Pfeffer, J., 2014. Social media for large studies of behavior. Science 346 (6213), 1063–1064. doi:10.1126/science.346.6213.1063. https://science.sciencemag.org/content/346/6213/1063

Sarsons, H., 2017. Interpreting signals in the labor market: evidence from medical referrals [Job Market Paper]. Working Paper 71. https://scholar.harvard.edu/sarsons/publications/interpreting-signals-evidence-medical-referrals

Similarweb (2022). Most Visited Websites - Top Websites Ranking for August 2022. https://perma.cc/D7VC-VU9H.

Widman, J., & Nicol, W. (2018). What is Reddit? A Beginner's Guide to the Front Page of the Internet. perma.cc/Z78A-U6ZJ.

Wu, A., 2017. Gender stereotyping in academia: evidence from economics job market rumors forum. IDEAS Working Paper Series from RePEc, 2017.

Young, A., Chaudhry, H.J., Pei, X., Arnhart, K., Dugan, M., Simons, K.B., 2021. FSMB Census of licensed physicians in the United States, 2020. Journal of Medical Regulation 107 (2), 57–64. doi:10.30770/2572-1852-107.2.57.