Crowdfunding in Health Care Consumption

Naomi Zewde and Nicholas Shengjun Wang[1]

Abstract

What is the role of crowdfunding in healthcare financing? Does it substitute for insurance coverage or provide complementary care? This paper evaluates the relationship between Medicaid expansion and the sentiment and frequency of medical fundraisers using data scraped from the Go Fund Me website. Difference-in-difference estimates indicate that expansion reduced campaigns tagged "medical" with no effect on placebos. Strongest and most precisely estimated results were in states with more pre-ACA uninsured. Campaigners' sentiment was unrelated to expansion but consistently more negative for medical fundraisers. Results suggest that crowdfunding acts as a substitute for coverage, stepping in when individuals are in need.

[1] Zewde: Graduate School of Public Health and Health Policy, City University of New York Naomi.Zewde@sph.cuny.edu; Wang: Union Bank of Switzerland sw3231@columbia.edu.

Crowdfunding, the process of soliciting financial contributions from one's social network, is popularly understood and academically regarded as a source of business capital (Calic 2018). Yet much of these funds are raised not for product development or creative endeavors but for medical care expenses. Rob Solomon, the CEO of popular crowdfunding site Go Fund Me, reported that one-third of all donations through the site were for healthcare expenditures (Cerullo 2019). With over 20 million Americans donating to over 120 thousand campaigns in a given year, Solomon was named by Time Magazine as one of the fifty most influential people in the healthcare sector (GoFundMe n.d.; Young 2020; Marty 2018). How do we characterize the role of crowdfunding in medical expenditures? Does crowdfunding constitute a component of the American system of healthcare financing?

Research on the crowdfunding sector is in its nascent stages. Among entrepreneurs, crowdfunding appears to act as a substitute for traditional capital markets. Entrepreneurs are able to find a better match with their potential investors, by connecting with individuals who have the highest willingness to pay for their particular niche product and are able to offer incentives to investors, like early access to products, that would be more difficult through traditional capital markets. Each of these characteristics position crowdfunding as a lower-cost substitute form of capital investment.

In medical care, crowdfunding may serve as a substitute for health insurance coverage. It is likely that individuals without insurance coverage will look to crowdfunding in place of the reimbursements otherwise provided by insurance. Charity care and the medical safety net system do not always offer access to the most reputable, highest volume, or highest quality healthcare facilities and therapies. For particularly advanced illnesses, for example, consumers may draw on crowdfunding for complex care outside of the scope of the safety net.

Alternatively, crowdfunding could serve as a complement to insurance coverage. Individuals who obtain coverage may receive additional information about their healthcare needs and may then be more likely to utilize healthcare and to seek assistance with obtaining further support for these (perhaps less urgent) medical needs. For example, one may become aware of the benefit of a knee replacement after gaining coverage or simply become able to obtain the surgery due to gaining coverage, and then turn to crowdfunding for the indirect costs associated with that major surgery or other out-of-pocket expenses not covered by insurance. Consumers often face substantial cost-sharing obligations, even with insurance coverage, and may seek further financial support after becoming insured. Finally, it is also possible that there is no observable relationship at all.

If crowdfunding serves as a substitute, increasing rates of insurance coverage may be associated with a decrease in the size of the crowdfunding sector. On the other hand, if crowdfunding serves as a gross complement, crowdfunding should continue to grow along with coverage and associated medical consumption, as individual fundraisers continue to seek support for more peripheral healthcare support not covered by insurance. Furthermore, if insurance and crowdfunding are complementary, then the kinds of care and the sentiment of the campaigner may observably vary with coverage. The pleas of an insured individual seeking complementary services may appear less dire than the campaign for the essential health benefits that insurance provides.

Our study examines the relationship between crowdfunding and the expansion of public healthcare coverage via the Medicaid program under the Affordable Care Act (ACA). We mine the website of Go Fund Me, one of the most popular crowdfunding websites (Kim 2018) and build a dataset containing the content and frequency of campaigns per month by state in the years spanning Medicaid expansion. We find a reduction in the rate of Go Fund Me campaigns tagged as "medical" following the expansion of Medicaid (-.21 per 100,000 low-income state residents uninsured prior to the ACA, $p < .01$), with no effect on placebo fundraising goals. The effect is concentrated in those states with the highest uninsured rates prior to Medicaid expansion.

Additionally, we construct sentiment scores of the mined campaign text to evaluate the writer's mood. Sentiment scoring has applied to mined text to predict electoral outcomes and to predict pricing and trading strategies in financial markets (Bollen, Mao, and Zeng 2011; Li et al. 2019; Zhang and Skiena 2010; Bermingham and Smeaton 2011). We found that the campaigners' sentiments were no less dire following expansion, further suggesting that medical crowdfunding addresses primary, and not complementary, health needs.

Results provide evidence that crowdfunding acts as a gross substitute for insurance coverage and in that sense comprises a de facto component of the American system of financing health care consumption. Crowdfunding may decline in the presence of popular proposed reforms to broaden access to direct healthcare financing, like a tax-financed single-payer system.

1. Data

Data Mining

We mine data from the text of user-generated crowdfunding campaigns on the Go Fund Me website. The aim is to collect a dataset of campaigns that identify three campaign features: when the campaign originated, where the campaign is located (which may differ from the poster's location), and what the campaign is for, whether for medical care or some other purpose. We collect these data by using the site's search function to search on these three campaign features, year Y, state S, and phrase P, which denotes the campaign's content. Our program completes a total of 3,500 searches, one for each year-state-phrase combination.

Set of searches:

$P :=$ {cancer,treatment,surgery,team,competition,film,art,music,school,student},

$Y :=$ {2013,2014,...,2018, 2019},

S:= {Alabama,Alaska,Arizona,...,Wisconsin,Wyoming}.

Each time, the searching target comprises a triple $t \in T$, where:

$$\text{T} := P \times Y \times S = \{(p,y,s) : p \in P, y \in Y, \text{ and } s \in S\}.$$

The terms included in the set of P phrases are drawn from campaigns listed under the website's browse feature. We scraped the text of all one hundred campaigns found via Go Fund Me's main web page from browsing under two medical categories (cancer and surgery) and three non-medical categories (education, competition, creative). We identified the most frequently used words in these campaigns' text to use as search terms. See Appendix A for word frequency maps and more on the search process.

At times, our intended content category differed from the tag we mined from the site. For example, a search for "team" may yield a campaign for help with a sports injury that the user has tagged as "medical" instead of "competition." We use the mined tag in our analyses, which were selected by the user during campaign setup. We conduct a robustness check dropping a portion of conflicting-tag campaigns and find our results were robust (see Appendix Table A2). We end with a dataset of 58,571 fundraiser observations, each with a single mined or imputed content category tag; mined location (city and state); and mined date of creation.

Sentiment Scoring

We assign sentiment scores indicating the sentiments expressed in the campaign text, after some initial word processing. We process the raw textual data from campaigns' fundraising stories using the Natural Language Toolkit (NLTK) in python. We normalize the textual data in the following order: a) replace contractions, b) convert to lowercase, c) remove numbers, and d) remove stop words such as articles and other short function words. We normalize word forms using lemmatization, which generates complete words from word stems, rather than stemming, which generates only word stems, to maximize the likelihood of finding a match within the word-sentiment database. Last, the remaining words are assigned a part of speech.

We adopt a lexicon-based approach to sentiment scoring that matches words from the mined campaign stories to SentiWordNet. SentiWordNet is a publicly-available database of words and their catalogued sentiments that has been shown to perform reliably in analyses of mined publicly-disclosed textual data (Baccianella, Esuli, and

Sebastiani 2010; Musto, Semeraro, and Polignano 2014).[2] The SentiWordNet database assigns each word in the English lexicon a three-dimensional score: a positive score, a negative score, and an objective score. The sum of the positive score and negative score equals the subjective score, and the sum of the subjective score and objective score is always equal to one. We use SentiWordNet to evaluate each word in the textual data and then calculate the average of the positive score, the negative score, and the objective score, respectively, for each campaign story. The overall sentiment for each story is calculated by taking the difference between the negative and positive sentiment scores. While we do not include the objective score in our inference tests, the results are robust to their inclusion. Additionally, we ran a robustness check removing the top and bottom 1% of sentiment scores (Appendix B). While again our results were robust to their exclusion, upon manual inspection, we identified textual irregularities among the most outlier scores including non-unicode characters. After exclusion of the 1%, 5%, and 10% of outliers on either end each produced similar a result, we ultimately selected the least restrictive exclusion.

---

[2] State-of-the-art sentiment analysis models fall within supervised machine learning settings, in which a set of texts are manually labeled or, more likely, are drawn from a well-labeled feature space used to train a theoretical model (Devlin et al. 2018; Yang et al. 2019). In reality, most data are unlabeled, as is the case with our crowdfunding text, and with a large volume of textual data, manually labeling sentiment is not infeasible.

We draw covariates from state-level data on insurance, medical spending, and population size from the State Health Access Data Assistance Center (SHADAC), which collates annual data from federal surveys including the National Health Interview Survey and the Current Population Survey to produce state-level aggregate information. We use state-year rates of "trouble paying medical bills," to proxy for healthcare affordability and unemployment rates to measure economic activity. For the size of the relevant population, we use the number of uninsured nonelderly adults (ages 19 to 64) with incomes below the threshold for Medicaid eligibility, 138% of the Federal Poverty Level (FPL).

Our primary outcome is the frequency of medical and non-medical campaigns occurring in a state before and after the state expands (if at all) Medicaid. To construct this measure, we first collapse our dataset on the sum of counts and mean of sentiment scores by state, month-in-year, and content category. After collapsing, we have 11,669 state-year-month observations across the four categories, as listed in Table 1. Go Fund Me began in 2010, and has few campaign records from that year. Nevertheless, our results were robust to the subset of data restricted to years 2013 and later (Appendix C). We then construct a campaign rate by scaling the count of campaigns by the size of the state's low-income uninsured population. We divide the state-year-month count by the number of low-income, pre-ACA uninsured, nonelderly adults in the state times 100,000.

The sentiment score is a measure of sentiment negativity. As such, the more positive the value, the worse the sentiment expressed. Medical fundraisers have the worst sentiment on average at -.010, while the remaining three categories are more similar to one another, between -0.027 and -0.032.

Table 1: Summary Statistics of Sample of Crowdfunding Campaigns

		Medical	Education	Creative	Competition	Total
Count	2010	1	2			3
	2011	25	7	6	7	45
	2012	30	42	8	15	95
	2013	470	445	131	212	1,258
	2014	1,730	1,196	354	584	3,864
	2015	4,840	2,898	714	1,198	9,650
	2016	5,420	3,481	903	1,327	11,131
	2017	5,364	3,289	1,006	1,361	11,020
	2018	5,427	3,023	944	1,290	10,684
	2019	5,673	2,908	1,045	1,195	10,821
	Total	28,980	17,291	5,111	7,189	58,571
State-Month-Years^{+}		3,727	3,335	2,058	2,549	11,669
Percent Post-Expansion		.564	.570	.626	.597	.584
Mean Campaign Rate*		3.061	2.317	.982	1.220	2.080
Std. Dev		3.450	4.516	1.873	2.082	3.448
Mean Sentiment Score		-.010	-.027	-.032	-.027	-.023
Std. Dev		.016	.016	.019	.019	.019

Note: Table 1 lists summary statistics of the set of Crowdfunding Campaigns scraped from the Go Fund Me website.

$^{+}$ State-Month-Years reflect the number of month-in-year observations across states.

* Campaign Rate is constructed as the count of monthly campaigns as a fraction of the number of nonelderly adult residents below the income threshold for Medicaid eligibility.

Our independent variable of interest is the interaction term "post-expansion," a binary indicator that equals one if the observation is both from an expansion state and from a year greater than the state's year of expansion, or equal to the year of expansion if implemented before June (Kaiser Family Foundation 2020). Just over half of the state-year-month observations are from the post-expansion period. Because of the staggered implementation of Medicaid expansion, with observations from at least one newly expanded state in each sample year, we include year-month fixed effects and state fixed effects in each model. In some specifications, we additionally include state-year unemployment rate and state-year rate of trouble paying medical bills drawn from SHADAC. Standard errors are clustered at the state level.

2. Empirical Analysis

Results from an event study analysis appear to satisfy the parallel trends assumption. The outcome is the rate of medical campaigns per 100,000 uninsured individuals below the Medicaid-eligibility income threshold prior to the ACA's passage. The analysis includes state and year-month fixed effects with clustered standard errors. We find clear reduction in the mean rate of campaigns for medical care following Medicaid expansion.

Figure 1: Event Study of Medicaid Expansion Effect on GoFundMe Medical Campaigns

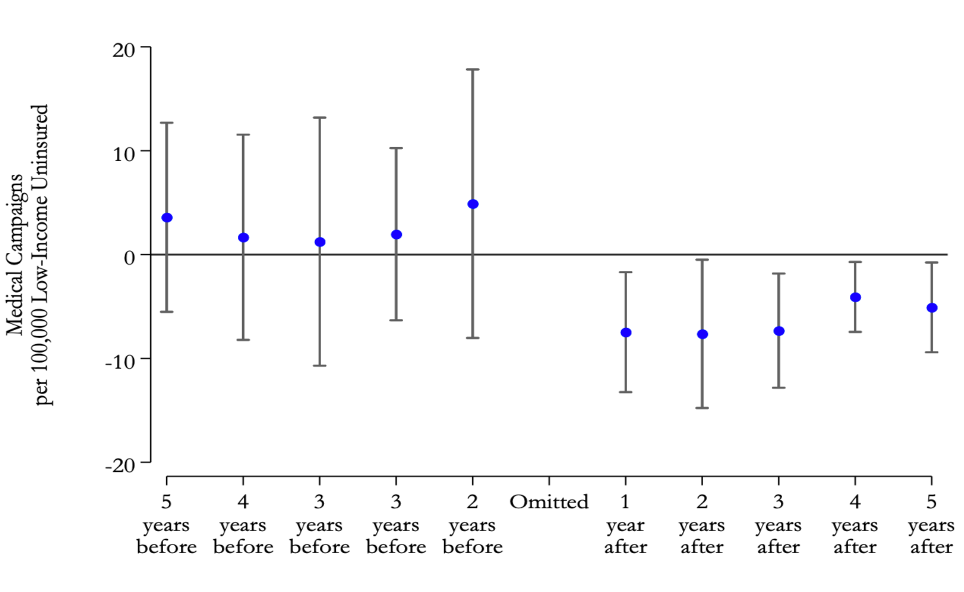


Table 2 presents results from a difference-in-difference analysis, including medical campaigns and non-medical placebo categories. We run poisson regressions on the rate of campaigns, stratified by content tag and including state and year-month fixed effects. We additionally run a set of regressions including the fixed effects as well as state-year-level covariates, labeled Model II. In both models, with and without covariates, the only significant relationship with Medicaid expansion is a reduction in medical campaigns.

Table 3 repeats the difference-in-difference analysis after stratifying by the number of pre-ACA uninsured low-income state residents. The relationship observed in our primary results, wherein Medicaid expansion is associated with a reduction in medical campaigns, is found to be concentrated within states that had the highest number of residents most likely to be directly affected by the implementation of Medicaid expansion.

Last, we examine the association between Medicaid expansion and the sentiment of campaigners' texts. We hypothesized that if medical campaigns persisted or increased despite the expansion of public coverage, we may find a concomitant improvement in the sentiments expressed by campaigners in their campaign narratives, which would suggest that the fundraising campaigns act as a complement to insurance coverage. Given that we did not find an increase in medical campaigns after expansion, and instead consistently found a decrease in their frequency, it follows from our earlier hypothesis that we would not expect to find an improvement in campaigners' sentiments.

Table 2: Expansion Impact on Rate of Go Fund Me Campaigns, Difference in Difference Estimate

	Model I				Model II			
	Medical	Education	Creative	Competition	Medical	Education	Creative	Competition
Post- Expansion	-.168	-.086	-.049	-.070	-.190	-.103	-.162	-.097
Std. Err	.061	.099	.081	.072	.077	.093	.103	.058
P-Value	.006	.388	.548	.330	.013	.268	.116	.097
State, Year-month FE	X	X	X	X	X	X	X	X
State covariates					X	X	X	X
Campaign Sentiment					X	X	X	X

Note: Data on crowdfunded campaigns retrieved from Go Fund Me website, totaling 58,571 campaigns across four categories, and 11,669 month-in-year observations across 50 states and four content categories. Robust standard errors are clustered by state. State-level covariates are: state unemployment rate, and percent of state residents reporting trouble paying medical bills, each by year.

Table 3: Expansion Impact on Rate of Go Fund Me Campaigns, Difference in Difference Stratified by Pre-ACA Uninsured

	Model I			Model II		
Pre-ACA Uninsured:	Low	Mid	High	Low	Mid	High
Post- Expansion	-.039	-.161	-.274	-.020	-.226	-.238
Std. Err	.070	.079	.043	.239	.131	.047
P-Value	.574	.041	.000	.932	.085	.000
FE	X	X	X	X	X	X
Trouble with Med Bills				X	X	X
Unemployment				X	X	X
Campaign sentiment				X	X	X

Note: Data on crowdfunded campaigns retrieved from Go Fund Me website, restricted to those tagged as medical. Data include a total of 28,980 campaigns, and 3,727 month-in-year observations across 50 states. Robust standard errors are clustered by state. State-level covariates are: state unemployment rate, and percent of state residents reporting trouble paying medical bills, each by year.

Table 4 presents results from a difference-in-difference OLS regression on sentiment scores for medical campaigns. As expected, we found no difference in the sentiments expressed in medical campaigns after Medicaid expansion, neither for the full sample nor after stratifying by the number of low-income pre-ACA uninsured state residents. We do find a significant and consistent relationship between average sentiment and the share of all monthly campaigns in a state that were tagged as medical. Additionally, in high-uninsured states, unemployment appears significantly and positively associated with campaign negativity, in that higher state unemployment is associated with more negative sentiment in the Go Fund Me plea, even after controlling for Medicaid expansion.

Table 4: Difference in Difference Estimate of the Impact of Medicaid Expansion on the Sentiment of Text in Go Fund Me Campaigns for Medical Care

		Total Population	Stratified by Pre-ACA Uninsured: Low	Mid	High
Post- Expansion		.002	-.002	.003	.002
	Std. Err	.001	.003	.002	.002
	P-Value	.157	.595	.078	.441
Share Medical		.022	.019	.021	.025
	Std. Err	.001	.003	.002	.002
	P-Value	.000	.000	.000	.000
Unemployment		.010	-.017	-.157	.244
	Std. Err	.065	.091	.104	.089
	P-Value	.884	.855	.151	.015
Trouble with Medical Bills		-.004	-.006	.024	-.014
	Std. Err	.007	.012	.016	.012
	P-Value	.517	.600	.169	.241
Fixed Effects					
	State	X	X	X	X
	Year-month	X	X	X	X

Note: Data on crowdfunded campaigns retrieved from Go Fund Me website, restricted to those tagged as medical. Data include a total of 28,980 campaigns, and 3,727 month-in-year observations across 50 states. Robust standard errors are clustered by state. Share Medical reflects the percent of all month-in-year campaigns within the state tagged as medical. Low Mid and High represent terciles of the number of state residents below the income threshold for Medicaid eligibility who were uninsured in 2010.

3. Discussion and Conclusion

Our results consistently provide evidence that medical crowdfunding is significantly related to a lack of insurance coverage and, in that sense, it fills part of a gap in the system of healthcare financing in the United States. Our event study depicted a decrease in medical crowdfunding in the years following the expansion of Medicaid, which was corroborated by difference in difference models. No other category of campaigns, including education, competition, and creative activities, exhibited a significant relationship. The significant relationship appeared only in states with the highest rates of pre-ACA uninsured low-income residents, those most affected by the public healthcare expansion.

Our study is limited by our inability to observe changes in individuals' insurance status. With individual-level insurance data, we could assess whether the reduction in medical crowdfunding reflects a reduction in the number of uninsured persons creating such campaigns. To address this limitation, we stratified by the state uninsured rate to isolate the population most likely to have undergone a change in health status. Still, we cannot distinguish between whether the association between state unemployment and negative campaign sentiment is driven by the individual's employment status, or their social network's. Finally, we do not have the universe of Go Fund Me campaigns, only the approximately 58,000 campaigns we collected through our systematic search. As a result, our estimates do not establish the overall size of the Go Fund Me medical campaign market, or the magnitude of its variation in response to insurance and employment.

In contrast to a cooperative complementary relationship, this study provides evidence that greater access to public healthcare coverage reduces individuals' need to campaign to social networks for financial contributions to cover the costs of healthcare. Our results raise the concern that episodes of economic depression may lead to higher rates of crowdfunded medical care, potentially with increasingly dire pleas as unemployment increases.

Finally, this phenomenon may be unique to the United States. Future research should compare the role of crowdfunding in the United States to that of other countries where healthcare financing relies less on the liquidity of consumers.

Appendix

Appendix A: Robustness to Multiple Search Phrase

We browse the Go Fund Me website under five available categories, 2 medical (cancer, surgery) and 3 non-medical (education, creative, and competition). We scrape the text of all one hundred campaigns associated with each browse category. After initial word processing on the scraped text using the Natural Language Tool Kit in python, we create word frequency maps to identify the most commonly used words in each category (Figures A1-A5). We select a total of ten search phrases among those frequently used in the search categories. See more at: https://github.com/swang2021/GoFundMe_scraper/

Appendix Table A1: Browse, Search, and Analytical Category Terms

Browse Category	Search Phrase	Analytical Category
Cancer Surgery	Cancer Surgery treatment	Medical
Education	School student	Education
Creative	Music Film Art	
Competition	Team Competition	Competition

Table A2: Summary Statistics of Sample of Crowdfunding Campaigns, Removing 25% Low Probability

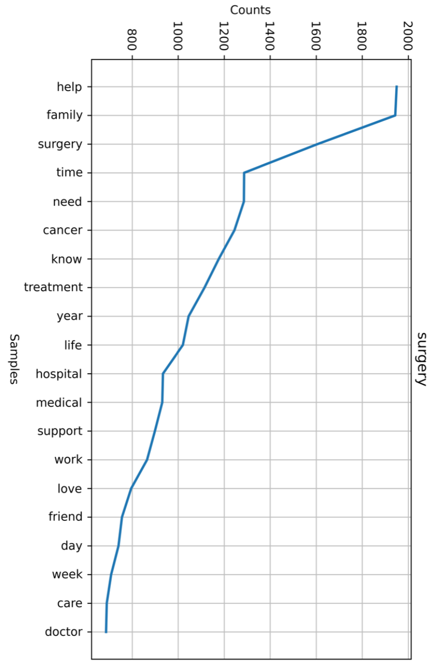
	Medical	Education	Creative	Competition	Total
Total Count	28,980	17,291	5,111	7,189	11,669
Percent Multiples	.287	.303	.288	.235	.280
Count After Drop	26,692	15,531	4,942	6,754	53,919
Mean Campaign Rate *	2.851	2.119	.958	1.180	1.946
Std. Dev	3.211	4.185	1.844	1.949	3.209
Mean Sentiment Score	-.010	-.028	-.032	-.027	-.023
Std. Dev	.016	.017	.019	.020	.020
Percent Post- Expansion	.566	.571	.626	.598	.585

* Campaign Rate constructed as state-year-month campaign count divided by number of nonelderly adult state residents with incomes below the Federal Poverty Level.

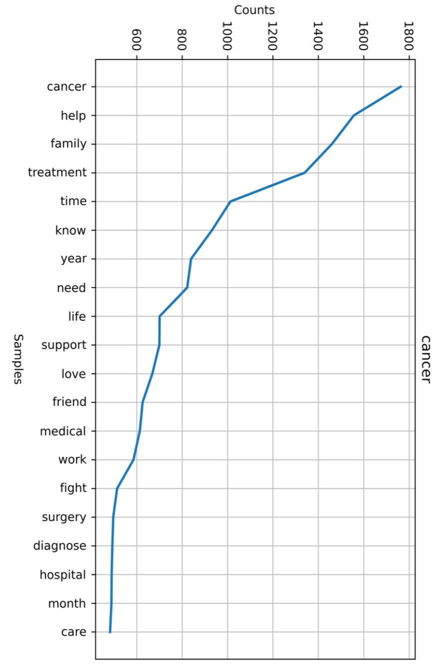
Table A3: GoFundMe Campaigns by Fundraiser Category, Dropping Lowest Probability Campaigns with Multiple Category Tags

	Model I				Model II			
	Medical	Education	Creative	Competition	Medical	Education	Creative	Competition
Post- Expansion	-.207	-.037	-.076	-.071	-.256	-.013	-.164	-.105
Std. Err	.069	.088	.070	.080	.071	.080	.107	.059
P-Value	.003	.677	.276	.372	.000	.873	.125	.076
State, Year-month FE	X	X	X	X	X	X	X	X
State covariates					X	X	X	X
Campaign Sentiment					X	X	X	X

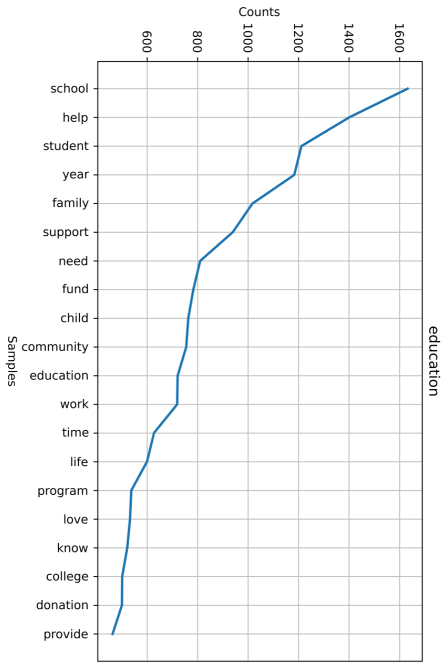
Appendix Figure A1: Word frequency map for campaigns tagged "Surgery"



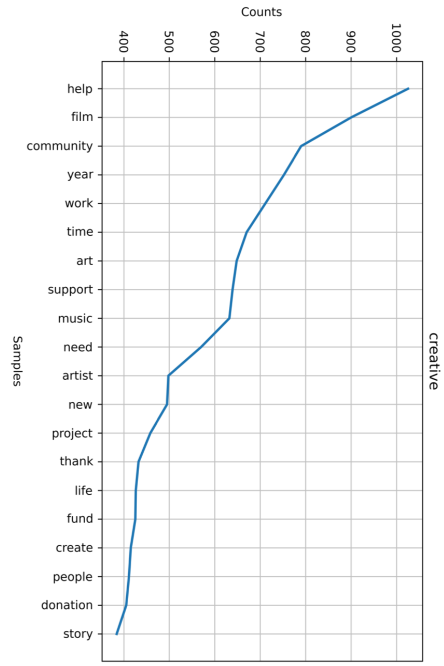
Appendix Figure A2: Word frequency map for campaigns tagged "Cancer"



Appendix Figure A3: Word frequency map for campaigns tagged "Education"



Appendix Figure A4: Word frequency map for campaigns tagged "Creative"



Appendix Figure A5: Word frequency map for campaigns tagged "Competition"

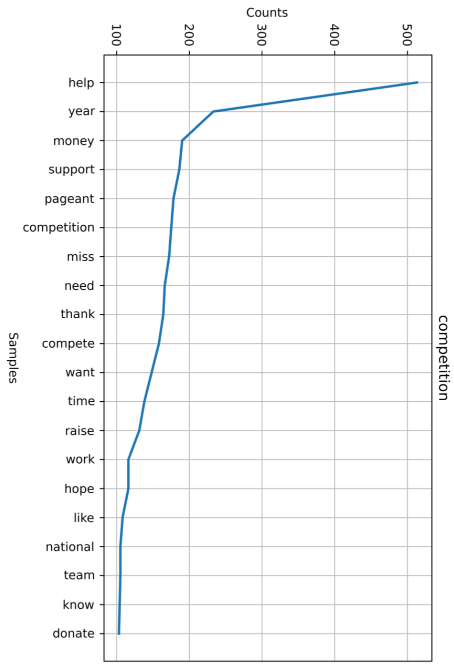


Table B1: Summary Statistics of Sample of Crowdfunding Campaigns, Dropping 1st and 100th percentiles of sentiment scores

	Medical	Education	Creative	Competition	Total
Total Count	28,980	17,291	5,111	7,189	11,669
Count After Drop	28,301	17,003	5,029	7,069	57,402
Mean Campaign Rate *	2.991	2.277	.973	1.214	2.044
Std. Dev	3.364	4.389	1.864	2.084	3.368
Mean Sentiment Score	-.011	-.027	-.031	-.026	-.022
Std. Dev	.014	.015	.018	.017	.018
Percent Post-Expansion	.565	.571	.627	.599	.585

* Campaign Rate constructed as state-year-month campaign count divided by number of nonelderly adult state residents with incomes below the Federal Poverty Level.

Table B2: GoFundMe Campaigns by Fundraiser Category, Dropping 1st and 100th percentiles of sentiment scores

	Model I				Model II			
	Medical	Education	Creative	Competition	Medical	Education	Creative	Competition
Post- Expansion	-.150	-.095	-.095	-.065	-.156	-.131	-.162	-.088
Std. Err	.056	.100	.067	.073	.072	.096	.111	.059
P-Value	.007	.343	.156	.379	.031	.175	.144	.136
State, Year-month FE	X	X	X	X	X	X	X	X
State covariates					X	X	X	X
Campaign Sentiment					X	X	X	X

Appendix C: Restricting to Years 2013 and Later

Appendix Table C1: Summary Statistics of Crowdfunding Campaigns, Restricting to Years 2013 and Later

	Medical	Education	Creative	Competition	Total
Total Count	28,980	17,291	5,111	7,189	11,669
Count After Drop	28,924	17,240	5,097	7,167	58,428
Mean Campaign Rate *	3.095	2.340	.987	1.228	2.097
Std. Dev	3.459	4.539	1.878	2.088	3.461
Mean Sentiment Score	-.010	-.027	-.032	-.027	-.023
Std. Dev	.016	.016	.019	.019	.019
Percent Post-Expansion	.571	.576	.628	.600	.589

* Campaign Rate constructed as state-year-month campaign count divided by number of nonelderly adult state residents with incomes below the Federal Poverty Level.

Table C2: Difference-in-Difference Estimates of Medicaid Expansion Effects on GoFundMe Campaigns by Fundraiser Category, Restricting to Years 2013 and Later

	Model I				Model II			
	Medical	Education	Creative	Competition	Medical	Education	Creative	Competition
Post- Expansion	-.166	-.088	-.049	-.071	-.199	-.107	-.162	-.098
Std. Err	.061	.100	.081	.072	.076	.093	.104	.058
P-Value	.006	.381	.548	.326	.009	.251	.118	.087
State, Year-month FE	X	X	X	X	X	X	X	X
State covariates					X	X	X	X
Campaign Sentiment					X	X	X	X