Fact or Fluff: Does Wording Used by Gene Editing Companies Affect Investor Behaviors?

Thomas Freireich

Dr. Michelle Connolly, Faculty Advisor

Honors thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University

Duke University

Durham, North Carolina

April 12, 2024

“One tiny drop changes everything.

At Theranos, we’re working to shape the future of lab testing.” - Theranos Website, 2014

Table of Contents

Acknowledgements

I would like to take this opportunity to extend my gratitude towards Dr. Michelle Connolly, who has helped me ideate, develop, and improve my thesis. I would also like to thank the members of Economics 496, who, on many occasions, provided valuable feedback to my research.

Abstract

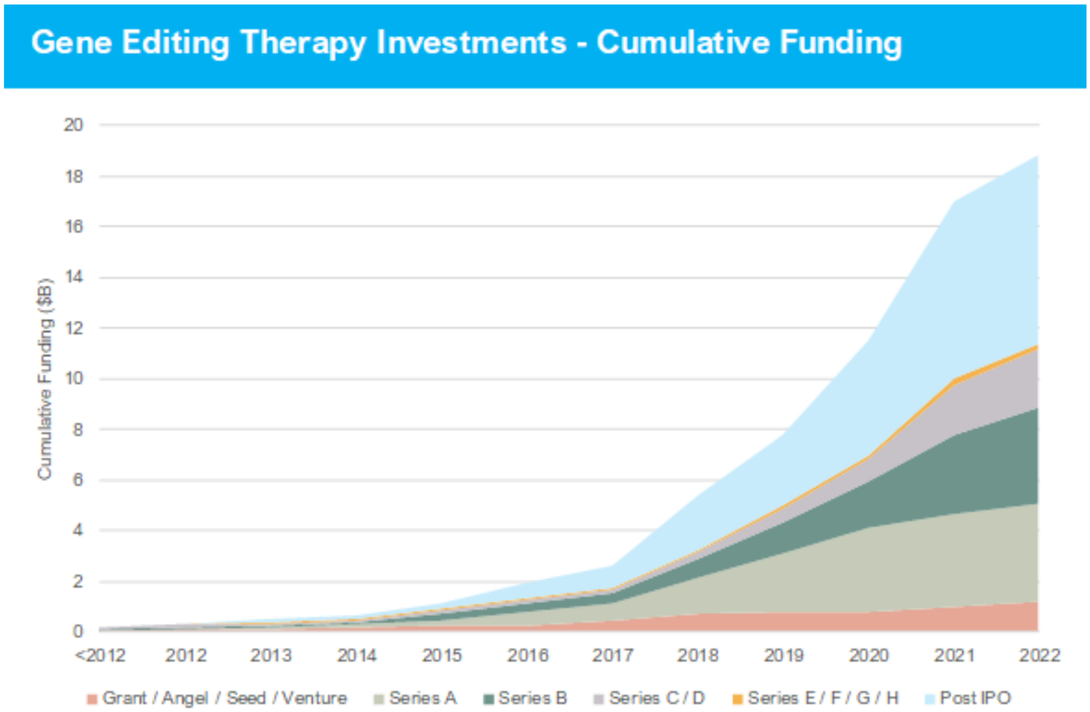
The writing style a startup uses to portray itself has an impact on investors' perceptions of them, subsequently affecting their venture capital decisions. This funding is particularly important given the prominence of venture capital as a primary financial source for growing early-stage biotechnology companies. Currently, due to recent scientific advances, many of these startup companies are utilizing novel gene editing based approaches to cure a variety of previously untreated diseases. For the sake of those affected, it is essential that this sector of the biotechnology industry is managed properly early on so that developed treatments can eventually reach FDA approval. This paper is in part inspired by recent happenings revolving around the fraudulent biotech startup, Theranos. Elizabeth Holmes, Theranos' founder, was renowned for making comments lauding the company's product. It seemed to many that investors were lulled by the idea of what Holmes made Theranos to be, invested in the company based on false verbal promises instead of the reality of the scientific product. Occurrences like the demise of Theranos are detrimental to both investors and competing companies in need of venture funds in order to develop their treatments. Thus, this paper explores the impact of word-usage and writing style on venture capital investment in various gene editing based startups, hoping to elucidate whether investors are being swayed by word choice.

Introduction

Within the biotechnological startup space, there exists a new and large sector of startup companies that aim to utilize new gene editing technologies to cure a variety of diseases. This novel technological approach, called gene therapy, is used to target genes associated with distinct diseases that induce some harm. This can be achieved through the insertion of specific properly functioning genes, the selective removal of certain genes, or both. This process is extremely novel, as it was not discovered until 2009. As a result, very minimal experimental testing of these therapies has occurred. In just 2018, the first experimental gene therapy treatment on a human took place (Gostimskaya, 2022). Despite its novelty, the technology itself has allowed for the emergence of a blossoming new branch of biotechnology, leading to many scientists' jumping at the opportunity to pursue cures for previously untreatable diseases.

As a result of the allure of gene editing to both scientists and investors, the market for gene therapies witnessed an immense growth, which occurred largely in part due to the promising outlook of various startups firms. From this, more and more scientists and investors have flocked to the gene-editing approach to disease treatment. Such a trend is observable in **Figure 1**, which demonstrates the drastic growth in the industry over the past decade as a measure of funding. This growth has continued through today, with many predicting a large-scale increase in the value of gene editing technologies in the future. Furthermore, while market valuations vary, many market analytics firms predict compounded annual growth rates (CAGRs) of 15 to 20% within the industry in the next ten years (Grand View Research).

Figure 1: A chart depicting the cumulative investment in gene editing firms from 2012-2022. Means of funding received are also shown by color (Emerson Insights).



As a result of the competition between these startup companies to bring a product has been likened to an "arms race," in which investment is paramount because it allows companies to optimally scale their research to develop their final drug product in as short a time span as legally feasible. Since these companies are not able to sell a product until several years of experimentation, clinical trials, and FDA approval, convincing investors to fund a zero-revenue company is quintessential for a startup gene editing firm to succeed. In fact, just recently, the first FDA approved gene editing therapy was completed in order to treat sickle cell anemia, with this treatment having a current value of $2.2 million (Lovelace and Kopf, 2023). This milestone drug approval has demonstrated both the feasibility of a gene-editing based product to be implemented into clinical practices to treat a previously untreatable and devastating disease while also revealing the true monetary value of such a treatment.

There is, of course, a large amount of risk associated with an investment in any startup, largely as it pertains to the product's in vivo (in organism) success. A lack of precedent by which to qualify investments in gene therapies has made the economic component of the industry very difficult to visualize, from initial investments to market entrance. However, given that venture capital investments are critical to the survival of individual firms, it is logical to conclude that companies will do whatever able to maximally accrue investment from these firms. Within this, it is to be expected that companies will be extremely selective in the language they use to portray itself and their product in order to maximally garner investment. **In a field in which the products can be unpredictable in terms of timetable and efficacy, this research will investigate whether the language that companies use can affect the amount of investment they receive from venture investors.**

II. Literature Review

The biotechnology industry encompasses all technologies that utilize live products to address a variety of diseases. The industry progresses concurrently with the development of new scientific products, and this progress, especially recently, has been largely driven by a startup culture that enables small companies to develop niche and new methodologies to address a specific problem. Today, the average biotech company is much younger and smaller in size than those 20+ years ago (Godfrey, Allen, Benson, 2020). Gene editing itself is a large portion of the developing biotech industry, with a projected market value of $36 billion in 2027 (Fletcher, 2023).

In the early stages of a biotech company, venture capital investment is necessary for a company to fund its initial drug development in its nascent stage since expenditures are very high, and profit is nonexistent. The other main feasible source of income, federal grants, are insignificant in value compared to any venture capital investment. As Ledley points out, the fact that gene therapies have not yet yielded market outcomes obscures the understanding of gene editing companies as economic firms (Ledley, McNamee, Udzil, 2014). The paper shows that while biotech early-stage investment is positively correlated to macroeconomic conditions, there is little relation between the state of the economy and the success of firms, which is measured by a firm's ability to survive, develop a product, and sustain profit or be bought out. Despite this unclear market, venture investments in gene editing have increased substantially since the early 2000s (Hanna, Remuzat, Auquiere, Toumi, 2017). This is likely because of both the proliferation of new companies and the fact that many companies entering clinical trials are yielding tangible results (Pian, Chandra, Stern, 2020).

Unsuccessful clinical results are detrimental to a firm's prolonged success, and word of such failures will more than likely deter investors. Success in experimental trials is the scientific component of a therapy that is crucial in achieving further investment. However, it turns out that, because of the extreme competitiveness that exists within the gene editing field, even in cases of positive experimental results (Godfrey, 2020), there can be startups that collapse. This demonstrates that there are a variety of factors aside from the science that impact investment approaches. As logically expected, lack of sufficient funding is one major cause of startup failure, (Ranjan, Jiang, 2015).

In general, venture capital investors across all industries have been shown to attribute great importance to revenue growth, international scalability, value added, and management team track record (Block Diegel, Fisch, 2023). However, in a gene editing context, many of these factors cannot be measured; revenue growth is nonexistent, international scalability is not a concern, value added is loosely defined by the market, and most management teams do not have a track record. Thus, it can be established that venture investing in gene editing startups is based upon a more unique set of parameters, as there is less financial data and prior similar investments for investors to reference. This implies that the use of certain language can carry increased weight in how investors view gene editing companies in comparison to companies in other industries or sectors.

In theory, investments should be made strictly on the scientific capabilities and market capabilities of a certain product; however, because of the uncertainty associated with scientific development, especially in early stages of company development, investors are often forced to look outside of the realm of strict scientific data to further determine their investments. Baum (Baum, Silverman, 2004) carried out research investigating the potential of various factors, such as upstream alliances, personal capital (i.e., quality of employees), and patents, to qualify venture capital investment, revealing that investors are motivated to invest in each company for both technological and personal reasons. Both high scores of personal and intellectual capital were correlated with more investment.

The aforementioned actions of Elizabeth Holmes and her former company, Theranos, provide a great illustration of the influence of marketing and word usage on potential investors. While Theranos claimed to be able to only need a drop of blood to conduct over 240 tests, this was not the case. Still, the company was able to rack up significant investments from venture capitalists and angel investors who were invested in Holmes' vision "to discover something new, something mankind didn't know was possible to do." After the exposing of the fact that her technology was ineffective, more than $600 million dollars were lost (Sheetz, 2018). In retrospect, investors, instead of investing in the scientific process behind a technology, invested in the idea of it.

In various contexts, word analysis has been used as a tool to quantify the language used by politicians, websites, entertainers, companies, and more. In the past, many analyses utilized a box of words model to quantify the usage of certain types of words in a text. Essentially, this model counts the frequency of words or a certain type of words in an excerpt. Adjectival frequency analysis is one form of this and has been performed previously in the context of analyzing advertisement behavior from firms (Ke, Wang, 2013) to determine the linguistic and rhetorical impact of adding superfluous content to a company's profile.

Moreover, recent developments in AI/ML have enabled the usage of natural language processing (NLP) analysis of words. In general, NLP involves the quantification of individual words by some measure to allow the AI to better understand the text. Sentiment analysis is one subset of NLP that quantifies words by their positivity, neutrality, and negativity. While words are assigned numeric values based on the emotional connotation of similar words, context is also considered. Sentiment analysis has been utilized in various contexts, from looking at consistency and bias in politicians' statements (Zaruba, 2021) to correlations between company news and stock price (Siering, 2012).

Research on the influence of word usage on venture investment is not an often-used form of analysis. Its investigation as a parameter occurs mainly in industries in which a firm's product and market fit are more defined, allowing profit to be an output. McLeod, Sears, and Chandler (2022) demonstrated that a logically based approach was relatively unsuccessful in garnering investment compared to an emotional or ethical appeal. Their study on word usage also revealed an association between word usage and company characteristics. Companies associated with less risk were observed to tend to appeal more to emotional and logical appeal, likely to build brand credibility and reliability. In contrast, companies with more inherent risk were shown to appeal more to logos, or a factual, numeric approach. The main takeaway from this research is that investors may be less receptive to factual information and more receptive to more emotional and personal matters of communication.

Studies in investments in crowdfunding campaigns (Bi, Liu, Usman, 2017) suggest that in the case where one is viewing an online profile and does not know much about a given product, words that resonate positively with an individual can affect investing decisions. This is significant in demonstrating that the verbiage used in digital profiles can be an effective means of communication from companies to customers. Still, the potential for application of such behaviors to professional investors' bevavios is unclear.

III. Data Overview

Due to the lack of availability of a dataset that focused on US-based gene editing companies, I created a dataset that recorded various characteristics of companies. This dataset had a total of 282 observations, each representing an investment in a firm at a given time. observations occurred between 150 firms, with a firm ID given to each firm that was tracked. I created several selection criteria for these firms. First, the firms had to receive some sort of venture capital investment. They also are required to have been founded after 2012, which marks the landmark commercialization of Cas/CRISPR gene editing related technologies. Firms were identified with online searches and databases like Crunchbase and Pitchbook. For each firm, utilizing LinkedIn and company websites, I identified several founder-related characteristics, including whether a company's founder has obtained a PhD in a STEM field, has attended a Top 20 School, per US News, or founded a prior biotechnology company. Another variable that was compiled for each firm was the sum of previous investment acquired from prior rounds of venture capital.

Figure 2: A table representing the percentages of founders of firms identified who have a PhD, attended a Top 20 School, and have founded a prior company.

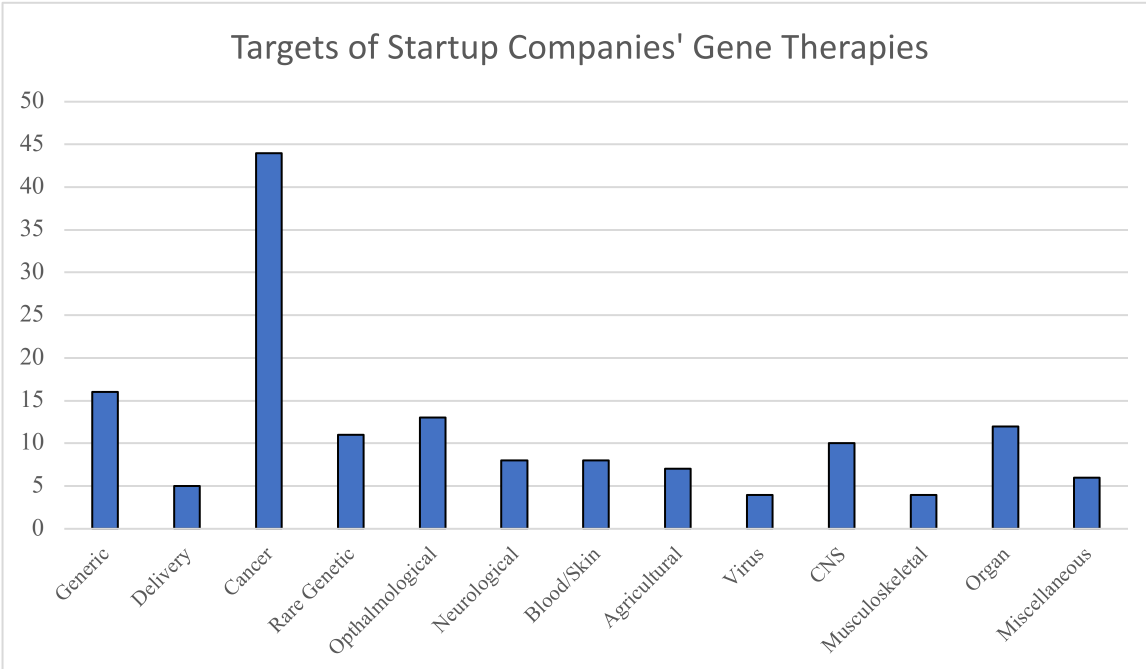
Founder Trait	Observations	Mean	SD
PhD	281	61.2%	48.8%
Top 20 School	281	68.0%	46.7%
Prior Company Founded	281	29.5%	45.7%

**Out of the 282 observations, one was dropped because of uncertainty in these measurements.*

From this, it seems that a large portion of founders for gene editing companies have PhDs in biological sciences. Similarly, many founders also seem to have attended a top 20 school. However, the 30% of founders having previously found a company stands out to me because it is much higher than I expected given the extensive developmental time span of biotech companies and unlikeliness that any founder has had previous experience founding a gene editing company. Firstly, because biotech startups experience a failure rate of about 90% (Reed, 2023), it is plausible that many of these founders have come from failed startups. Additionally, it is also plausible that only the few who were able to successfully sustain a startup in the past are able to do such again. Still, the fact that 30% of founders of gene editing firms that have received venture funding suggests that experience in the biotech industry may be valuable to investors.

Additionally, for each firm, the branch of disease that their drug products target was tracked and categorized. These "genres" of treatments include 1) generic, 2) drug delivery, 3) oncological, 4) rare genetic disease, 5) ophthalmological, 6) neurological, 7) blood/skin related, 8) agricultural, 9) virological, 10) central nervous system (CNS), 11) musculoskeletal, 12) organ related, and 13) miscellaneous. In rare cases in which a company had two therapies that targeted different diseases, the product farthest along their clinical pipeline at the time of investment decided the assigned category.

Figure 3: A bar graph depicting the number of firms targeting certain types of disease



**Out of the 282 measurements, 4 were dropped due to uncertainty in their disease category*

Figure 3 demonstrates the variability between the treatments that firms are pursuing, as evidenced by the consistency in every category of disease target except cancer. However, cancer is likely an outlier because of both a lack of current effective treatments and the variety that exists within it, likely increasing investor appeal. There are over 200 types of cancer, and they can all arise differentially depending on their underlying cause. Moreover, large populations are affected with various forms of cancers. On the other hand, while neurological disorders are prevalent, there only exist a few, such as Alzheimer's, Parkinson's, and Huntington's Diseases, that affect large populations. Or, for virological treatments, excluding a few rare cases, vaccines have been effective in treatment, limiting the opportunity for gene editing companies to pursue viral infections.

The dependent variable was measured in investment dollars ($) in real terms of 2023, with each round and investment year being categorized for each company. Sums of investments before a given round were also tracked for each individual firm. Moreover, only companies whose portfolios of products were within a single medical category were selected. For instance, some companies target ophthalmological disorders, and thus investor appeal pertains to this, and the investment can be accounted to investment in a certain market. If a company had products in two different types of therapy, it would be unclear for which product the investment was made. Thus, I intend to utilize sentiment analysis to generate a quantitative input for

Given that the major source of funding for biotechnology startups is from venture capital investments, I wanted to evaluate this investment as an output for a series of some independent variables. Venture investments can be made at various points in a startup company's timeline, and typically occur in the order of pre-seed, seed, series A, series B, series C, up to series F. Many firms do not choose to partake in all of these rounds of funding, instead picking when to undergo a round of funding based on a variety of factors. Typically, investments in earlier rounds have lower monetary value and more equity stake as a result of the increased risk in investment. In contrast, investments from later rounds are often greater in monetary value but lower in equity. This idea is supported by **Figure 4**, which entails a summary of the values of investments in given rounds. While a steady increase in investment is observed as rounds progress, large standard deviations seem to suggest inconsistency between investments in the same round, as observed with the Series A investments in Allogene and Apriori Bio being $498 million and $0.052 million, respectively.

Figure 4: Summary statistics of investments in gene editing startups, by round

Round	Observations	Mean	SD
Seed	38	18.9	46.5
Series A	107	77.3	107.2
Series B	90	100.2	62.8
Series C	47	127.5	105.56

While many studies analyze venture investment as a measure of company survival or proliferation, Hand (2007) takes a unique approach in utilizing venture capital funding as a dependent variable by which to evaluate round to round returns for US biotech companies. He supports the usage of venture funding as an output because of its encapsulation of a variety of endogenous characteristics of a firm, such as its size and predicted growth. It is important to recognize that Hand largely introduces this metric because of the industry driven context of the research; biotech startups often have few short-term indicators of success, and valuations can be inconsistent. Thus, I intend to use venture capital funding as my dependent variable assessing company output as a result of my established independent variables.

To quantify verbiage, I first accessed firms' websites six months prior to an investment (or as close as possible) utilizing Google's Wayback Machine database. Then, I captured the text on the company's websites' main screen and entered it into SpeakAI's sentimental analysis in order to assess the language of this input. In addition, I used an adjective analysis tool to determine the percentage of adjectives in text, chose the company's websites because, while private pitches are not accessible, startup biotech companies are very intentional about the verbiage they use, and the home page of a website provides a firm the opportunity to craft an image for itself. Because of the freedom in terms of what a firm can do with their websites, they are ideal sources of wording compared to other potential publicly accessible options such as federal filings that are much more formal and standardized in nature. Moreover, because biotech startups are so intentional and selective in their wording, it is very likely that the style of writing seen on a company's website is representative of how they talk about themselves in more investor-specific contexts.

SpeakAI's sentiment analysis tool analyzes the text and provides outputs in terms of positivity, negativity, and neutrality. Additionally, I developed an "emotion" variable that is the sum of the positive and negative components of the text.

Figure 5: Statistical summaries of sentiment analysis variables

	Observations	**Mean**	**SD**
Positivity %	282	46.6%	31.7%
Neutrality %	282	35.6%	30.6%
Negativity %	282	17.8%	26.3%
Emotion %	282	64.4%	30.6%
Adjective %	282	12.1%	5.2%

Figure 5 demonstrates that firms tend to use positive language while avoiding negativity, which is expected and indicative of the fact that the sentiment analysis tool is likely effective. One concern, however, regarding SpeakAI's software was the fact that it read certain small phrases incorrectly (Collins, Johnston, 2023). For instance, the tool deemed the statement "curing cancer is our mission," to be 100% negative, and "we aim to prevent cures" to be 100% positive. Moreover, I noticed that many of the excerpts with fewer words tend to give sentiment values at certain major intervals, such as 25%, 33.3%, 50%, and 100%. To me, this is suggestive of the fact that the tool does struggle with passages consisting of fewer words, as this limits

context and diminishes the ability of the AI to optimally understand the text. To address this, I ensured that excerpts had to be at least 75 words in length. As Collins and Johnston note, short phrases seem to be an exception to the generally accurate tool. For supplementary context, I have provided several examples of texts and the scores that they earned for positivity, negativity, and neutrality to the appendix.

Because the methodology that SpeakAI uses to describe its algorithm is confidential, I decided to carry out a regression between my various points to determine if the positivity, negativity, and neutrality outputs might be affected by my adjective count. From this, I determined that more neutrality was significantly associated with more adjectives at a 5% level. Furthermore, more negativity was found to be significantly related to greater adjective counts. It is likely that these results could be a result of the common adjectives used in the type of passages used to generate the data points or just a byproduct of human language. Regardless, the fact that these values both follow expectations and are not varied to an extreme extent further suggest that the SpeakAI tool is acceptable to use.

Figure 6: Regressions of SpeakAI sentiment analysis and adjective count

VARIABLES	Adjective Count
Positivity	0.00439
	(0.00981)
Constant	0.119***
	(0.00552)
Observations	286
R-squared	0.001

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	Adjective Count
Negativity	0.0225*
	(0.0118)
Constant	0.117***
	(0.00371)
Observations	286
R-squared	0.013

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	Adjective Count
Neutrality	-0.0209**
	(0.0101)
Constant	0.129***
	(0.00471)
Observations	286
R-squared	0.015

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

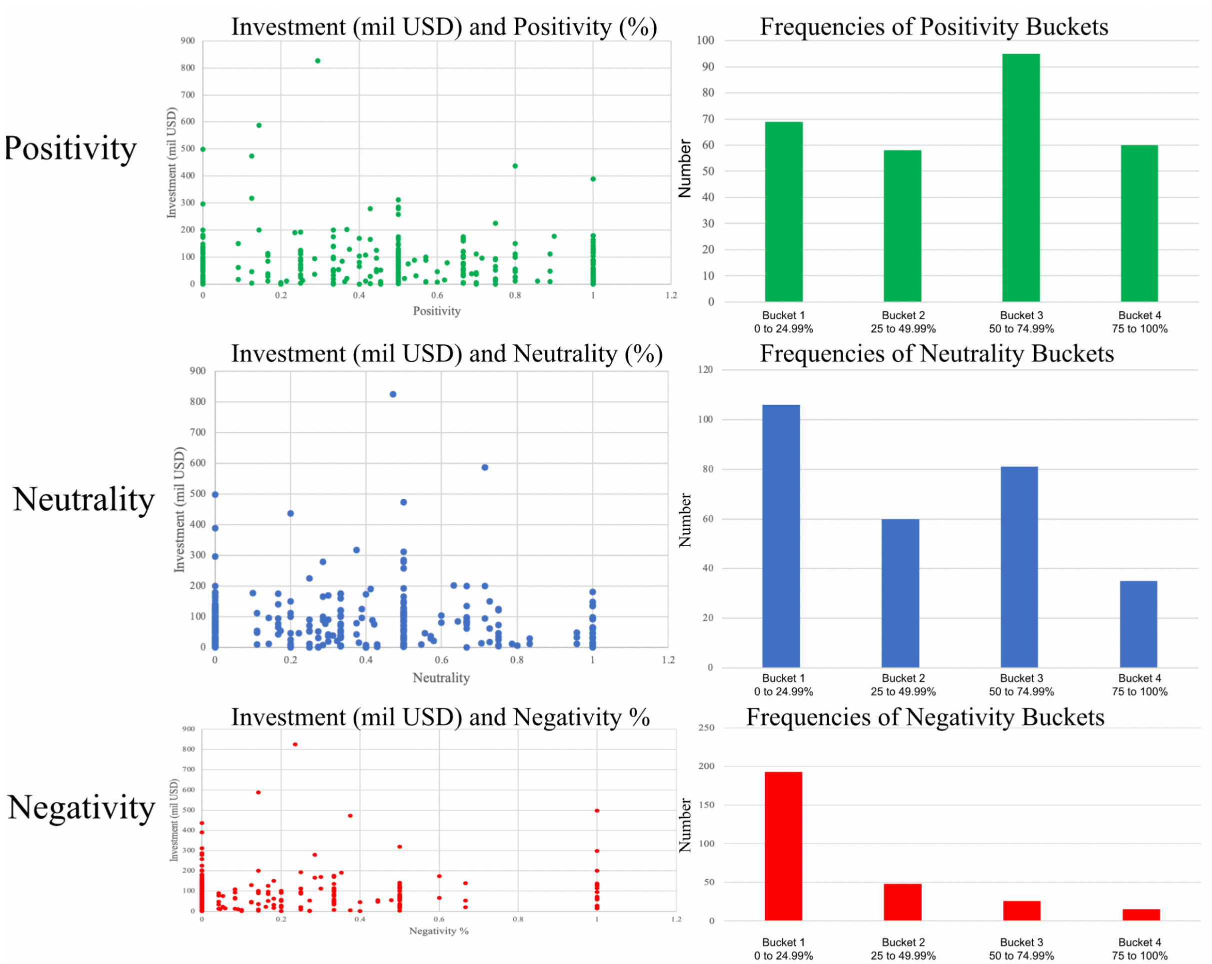
IV. Empirical Specification

It is important to acknowledge the lack of reference regarding specific investments into gene editing. The research behind the technology occurred in the early 2000s, with the first official company using the technology CRISPR therapeutics founded in 2012. Because the industry is very limited in both scientific markers, financial analyses, and public availability, the few analyses regarding gene editing investments largely look at patent rights and founder traits (Baum, 2004). Word usage, however, has been shown to sway investor opinions (Taffler, 2017), and will for the first time be considered as a factor gene editing startups use to attract venture capitalists. To perform my analysis, I run equation (1).

$$(1)\ Investment_{it} = \beta_0 + \beta_1 Language_{it} + \beta_2 FounderTraits_i + \beta_3 PreviousFunding_i + \beta_4 YearEffect + \beta_5 TherapyTargetEffect + \beta_6 FirmEffect + \beta_7 RoundEffect + \varepsilon_{it}$$

In the above equation, $Investment_{it}$ represents the investment, in million USD dollars, in a given round for a company. There are a variety of factors that have been shown to have some impact in venture capital investment in biotechnology. Thus, I control for as many potentially confounding variables as possible by introducing various terms that aid in the isolation of language as the main independent variable driving investment. Here, β_0 represents the average investment in a benchmark case. β_1 represents the coefficient to the variable tracking a certain form of language analysis. In addition to running regressions with both the numerical outputs from the sentiment analyzer as *Language* values, I decided to categorize the sentiment analyzer percentages into quartile sized buckets for each form of language and run a fixed regressions with these variables. The introduction of buckets of words allows me to analyze whether the effect of certain extents of sentiments is significant. Furthermore, it helps mitigate the earlier mentioned problems of repeated percentage values, as depicted in **Figure** 7. For instance, if bucket 4 for positivity consists of text excerpts that are very positive, I could establish whether extreme positivity is correlated to investment. This could account for inconsistencies of the numeric approach and the apparent lack of specificity of the SpeakAI tool. For these regressions, my benchmark was bucket 1, which consists of all data points in which a specific sentiment's percentage score is 0 to 24.99%.

Figure 7: Scatter plots representing the repetition of certain positivity, negativity, and neutrality percentages and how categorization affected data distribution



In summary, the forms of *Language* with which I will run the regression in equation (1) are positivity, positivity buckets, negativity, negativity buckets, neutrality, neutrality buckets, emotion (positivity + negativity), emotion buckets, and adjective count. Emotion was included as a variable to account for my hesitation regarding the efficacy of the SpeakAI model in deciphering positivity and negativity; instead, this metric captures the overall emotional appeal of a chosen excerpt. I included adjective counts because greater adjectival use is often correlated with superfluous wording that firms could be using to try to influence investors. The slope of β_1 and its associated error should reveal the significance of the effect of the type of verbiage a biotech firm uses and the venture investment it receives. Should the values be non-significant or significant but minimal in value, this would imply that the type of verbiage a firm is using to create its image is not affecting venture investment. On the other hand, significant values for β_1 could imply that venture investors in gene editing are being swayed by the verbiage a startup is using.

FounderTraits serves as a collection of dummy variables that represent whether a founder of a firm has received a PhD, founded a previous biotech company, and attended a top 20 school. I included these variables because of the allure that they may have in acquiring investors. Whether it is the research experience of a PhD, the industry knowledge of a previous founder, or the prestige of having attended a top institution, it is very fathomable that investors may look favorably upon companies founded by individuals with these characteristics in the early stages of a gene editing startup. β_2 is the coefficient associated with these traits, and regression analysis should reveal the individual contributions of the three recorded components of founder value. β_3 is the coefficient for *PreviousFunding*, which I included as a numerical value in mil USD, because it is likely that companies that received sizable venture capital investments in prior rounds are more well positioned to have a greater investment in a later round, in comparison to a company that did not have prior funding.

The remaining variables of my regression, including year of investment, therapeutic target, firm ID, venture investment round, are all categorical in nature and are associated respectively with β_4, β_5, β_6, and β_7. The year of investment was included to account for potential macroeconomic fluctuations that could result in changes in investment behavior. Yearly variations, such as hype and macroeconomic conditions, leading to differential changes in investment in gene editing firms such as hype, should be captured within β_4. β_5, the coefficient associated with the category of therapeutic target (page 13) should relate the effect of the market for a certain drug to its investment value. It is incredibly difficult to evaluate the market appeal of certain treatments, as gene editing treatments have no existent pricing model coefficient to the expected market value of a drug. One approach that has been utilized previously is the total accessible market (TAM) approach which looks at the population affected by a disease; however, this approach is flawed in that it does not look at the severity of diseases, the accessibility of treatments, and the presence of alternate treatments and the value of these options. Thus, I categorized disease targets to allow β_5 to inherently absorb these variations between very different treatments.

Moreover, given that there are several firms that appear twice in my dataset due to their receiving multiple investments over the course of their growth, I included a firm effect that should account for fluctuations that occur as a result of the unique behaviors of a firm. Thus, β_6 is the best way to encapsulate the effects of the individuality of a firm on the investment it receives. Also, the round in which an investment occurs has a significant impact on the absolute value of funding received. As previously mentioned, earlier round investments often consist of significantly less money due to high associated risks, and later round investments typically increase in terms of value because there is less risk associated with the firm's product. Thus, β_7 should hope to account for round-to-round variations that could affect an investment at a given time point. The error in the function is accounted for with the error term.

For these regressions, I determined that it was optimal for my regression's fixed effects to be done with a random effects regression because of the large omitted variable bias that is likely occurring. Additionally, because there is little association between individual specific effects and the language that a company uses, as evidenced by the lack statistically significant difference between the correlations of my fixed effect and random effect models for Positivity, Negativity, Neutrality, Emotion Adjective count. Also expectedly, a random effects model provides the added benefit of greater R^2 values because this approach accounts for random variability that occurs between firms and derive a maximally precise regression analysis. For the *YearEffect*, I chose the year 2019 due to the fact that this was the year with the highest number of investments before COVID, allowing me to compare post-COVID and pre-COVID investment trends. For the *TherapyTargetEffect*, I chose my benchmark to be generic drugs, as this should enable an effective evaluation of what types of therapies may receive more or less investment in comparison to firms developing generic gene editing tools. I selected my benchmark round as Series A, because although it occurs sequentially after the Seed round, it has more observations than other rounds and can serve as an intermediary between earlier rounds that typically receive lower funding and later rounds that typically receive more.

V. Results

First, I carried out the regression analysis for the numerical sentiment values, assessing correlation between positivity, neutrality, negativity, emotion, and adjective count. As observed in **figure 8**, we can see that the correlation between sentiment value and investment is not statistically significant. Still, the magnitude of the coefficients can still hold some weight. For instance, it seems that negativity is positively correlated to investment in a startup. Specifically, the model predicts that a 1% increase in negativity is associated with a $415,500 dollar increase in investment. While this is not necessarily accurate in all cases, this result is surprising, as I would expect negativity to potentially hurt a company's investment prospects. However, it is plausible that investors who see a company to be maintaining transparency and authenticity. Moreover, this could be a result of companies utilizing a more scientific approach that invokes words often associated with negativity. For instance, if a company chooses to mention the significance and debilitating nature of a disease, this may increase the negativity score they received while potentially increasing investor appeal.

Additionally, the fact that the positivity score was negative in its correlation to investment was also surprising to me and suggests that, while the results are not statically significant, there may be a general avoidance of companies that are overly positive. The model determined that a 1% increase in positivity is associated with a $272,200 decrease in investment. It is plausible that investors, however, are wary of products that utilize overly positive language and less disease specific verbiage, which would likely be more associated with a negative score. The correlations to neutrality, emotion, and adjective count suggest that these variables are having a minimal impact on investment, as demonstrated by their values being lesser in absolute value and having a minimal impact on investment, as demonstrated by their values being lesser in absolute value and having confidence intervals that include both positive and negative values. Still, it is important to exercise caution in the analysis of such variables due to their lack of statistical significance. Also, adjective count seemed to have no significant impact on investment. This indicates that investors are not affected by writing styles that tend to have more fluff or unnecessary verbiage. However, it is important to recognize the limitation of sheerly counting adjective percentages in that this does not effectively capture the differences in specific adjectives used and context.

Figure 8: Investment (million USD) Regressed with Sentiment %

	Positivity %	Neutrality %	Negativity %	Emotion %	Adjective %
VARIABLES	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)
Language	-27.72	7.649	41.55	-7.716	-11.46
	(26.75)	(27.61)	(38.14)	(27.63)	(159.3)
PhD	88.23*	86.17	89.80*	86.17	85.28
	(52.48)	(52.68)	(52.52)	(52.68)	(54.50)
Top 20 School	-145.6	-145.5	-167.9*	-145.5	-149.4
	(92.86)	(94.59)	(94.16)	(94.58)	(93.70)
Prior Company Found	321.2**	311.8**	333.3**	311.8**	312.2**
	(141.2)	(141.7)	(142.1)	(141.7)	(142.6)
Prior Funding (mil USD)	0.117	0.116	0.125	0.116	0.120
	(0.171)	(0.172)	(0.171)	(0.172)	(0.177)
	Positivity	Neutrality	Negativity	Emotion	Adjective %
TARGET EFFECT	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)
Drug Delivery	-167.5*	-164.2	-197.3*	-164.2	-168.4*
	(98.68)	(101.2)	(101.7)	(101.2)	(101.4)
Cancer	211.9	201.1	222.6	201.1	201.2
	(153.5)	(153.9)	(154.3)	(153.9)	(154.6)
Rare Genetic Disease	77.89	67.60	51.63	67.66	62.72
	(80.40)	(82.39)	(79.23)	(82.39)	(81.76)
Ophthalmological	174.2	164.8	187.0	164.8	165.8
	(134.9)	(135.4)	(136.0)	(135.4)	(135.6)
Neurological	-160.4**	-150.5**	-144.4**	-150.6**	-145.6**
	(74.29)	(74.93)	(73.01)	(74.94)	(74.09)
Blood/Skin	219.4*	218.3*	217.5*	218.3*	218.1
	(132.0)	(132.6)	(131.9)	(132.6)	(132.7)
Agricultural	-208.0**	-219.9**	-242.1***	-219.9**	-226.7**
	(92.53)	(95.50)	(91.35)	(95.50)	(93.01)
Virological	333.3***	316.1***	317.5***	316.1***	310.6***
	(92.08)	(91.71)	(89.71)	(91.71)	(90.32)
CNS	216.1**	204.6**	207.8**	204.7**	201.0**
	(95.00)	(94.96)	(94.11)	(94.97)	(95.11)
Musculoskeletal	257.6*	251.8*	264.6*	251.8*	251.3*
	(147.2)	(147.8)	(147.4)	(147.8)	(148.9)
Organs	278.6	256.8	278.6	256.9	252.6
	(178.6)	(178.0)	(178.3)	(178.0)	(180.2)

Miscellaneous	-92.92	-93.74	-112.6	-93.71	-97.27
	(73.52)	(75.14)	(74.57)	(75.14)	(74.11)
	Positivity	Neutrality	Negativity	Emotion	Adjective %
ROUND EFFECT	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)
Seed Round	-70.02***	-70.99***	-70.73***	-70.99***	-70.97***
	(17.40)	(17.46)	(17.36)	(17.46)	(17.83)
Series B	31.41*	31.15*	31.44*	31.15*	31.05*
	(16.81)	(16.88)	(16.80)	(16.88)	(16.92)
Series C	64.00**	63.58**	63.03**	63.57**	63.05**
	(31.15)	(31.29)	(31.13)	(31.29)	(31.54)
	Positivity	Neutrality	Negativity	Emotion	Adjective %
YEAR EFFECT	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)
2013	0.677	4.453	9.278	4.436	6.621
	(50.63)	(51.11)	(50.34)	(51.11)	(50.55)
2014	-6.108	-5.506	-0.365	-5.514	-4.207
	(40.88)	(41.26)	(40.98)	(41.26)	(41.05)
2015	-6.386	-5.169	-3.095	-5.167	-4.152
	(32.33)	(32.54)	(32.28)	(32.54)	(32.56)
2016	-39.79	-40.06	-36.00	-40.07	-38.93
	(29.53)	(29.77)	(29.67)	(29.77)	(30.18)
2017	-15.32	-15.76	-11.33	-15.77	-14.60
	(25.56)	(25.82)	(25.76)	(25.82)	(26.28)
2018	-21.79	-23.36	-19.75	-23.36	-22.78
	(21.86)	(21.94)	(22.03)	(21.94)	(22.39)
2020	-26.37	-26.14	-23.67	-26.14	-25.41
	(22.51)	(22.69)	(22.55)	(22.69)	(22.71)
2021	23.66	23.43	26.27	23.43	24.15
	(24.32)	(24.50)	(24.40)	(24.50)	(24.61)
2022	26.36	26.74	29.27	26.73	27.72
	(33.47)	(33.68)	(33.48)	(33.68)	(33.96)
2023	-4.382	-3.245	0.719	-3.257	-2.003
	(31.62)	(31.95)	(31.63)	(31.95)	(31.72)
2024	-43.51	-51.80	-58.25	-51.77	-55.54
	(89.06)	(89.74)	(88.27)	(89.74)	(88.71)
Constant	-84.55	-90.39	-92.58	-82.73	-82.51
	(84.36)	(87.64)	(84.67)	(84.89)	(87.90)
Adjusted R^2	.8349	.8335	.8351	.8335	.8334
Observations	277	277	277	277	277
Number of Round	4	4	4	4	4

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The analysis of founder value reveals that the presence of a PhD was significant at a 10% level in the regressions measuring positive and negative language, with founders having a PhD associated with an 85-90 million USD increase in investment. These results adhered to my expectations, as I would expect a founder having a PhD in biological sciences to increase the credibility and pedigree of a certain company because it signifies that a founder is more likely to be scientifically familiar with the company's research. However, it seems that, surprisingly, investors were not as favorable towards those who attended top 20 schools, with this trait associated with 145-170 million USD decreases in value. Still, these results do lack statistical significance in the regressions not using negativity and cannot ensure such a relationship.

However, it seems that having a founder who had previously founded a company significantly marks the greatest increase in investment value. Depending on which sentiment the model was assessing, the value of having founded a prior company ranged from about 310-330 million USD. It seems as though investors have been placing extreme value on industry experience. Moreover, it is very likely that the 30% of companies founded by a founder with previous experience likely were some of the few founders to previously have success before. In the biotech startup space, it is very common for startups to be bought out by larger pharmaceutical companies, so it is feasible that founders of previously bought out companies have decided to enter the new space of gene editing. Additionally, there are several prominent figures in gene editing whose association with the founding of a company (even if they are uninvolved) could boost a startup's apparent value. So, perhaps this variable encompasses a networking effect in which more connected individuals are able to not only achieve credibility by their experience alone but also by support from prominent scientific names who are well connected.

Interestingly, it seems that previous funding is not largely an indicator of the sort of investment that a firm receives. This could be a result of the firm fixed effect, which may more capture firm to firm variation more effectively than the previous funding amount. It could also be that this factor is not a significant consideration to investors because there is little carryover in terms of who will invest in a firm from one round to the next or simply the everchanging nature of a firm's scientific progress.

From this regression, the fact that the constants are not statistically significant at a 10% level indicates some level of concern regarding the effectiveness of the model. This occurrence could be due to omitted variable bias. The nature of the gene editing sector is such that there is minimal available data with which to qualify an investment. Not only are many details specific to the investment itself confidential, but it is also incredibly difficult to ascertain the value of certain treatments. Moreover, as previously mentioned, the fact that there are limited financial indicators of success could be limiting my model. Additionally, the constant being negative is somewhat of a concern because the benchmark case should receive a funding amount above zero; however, it is likely that the same contributors to the lack of significance in the constant are causing this. Furthermore, the value of a Prior Founder being so high is likely contributing to this negative value greatly. Interestingly, I did notice that the simplified versions of my model excluding certain terms seemed to increase the significance of my constants while decreasing my R^2 values, suggesting that my inclusion of more variables should absorb what effects were previously associated with the constant. I also checked the correlation matrix for my explanatory variables and did not observe significant instances of multicollinearity.

While the R^2 values are promisingly high, suggesting that my regression has reasonable predictability. Moreover, while my "between" R^2 values, which represents the relationship of subgroups of my dataset, are 1.000 due to the random effect model picking up random noise and including it in the analysis, my "within" R^2 values, which represent the relationship between my subgroups and investment, resides at around 0.8, substantiating the idea that the model could be used an effective predictor. These values were much higher than expected.

Results in **Figure 8** also demonstrate the relationship between the various therapeutics target categories and investment received in the non-categorical regressions. Between different regression inputs for language, the general trends and significance of therapies targeting a certain type of disease seemed to be consistent. Significance in these results is beneficial because it could enable a deeper understanding of the drivers of drug value. If investment in a certain type of treatment is either positively or negatively significant, it would suggest that investors are willing to spend more money on a given type of treatment because of what it targets. Here, drug delivery is significantly negative at the 10% level for positivity, negativity, and adjective count. It was surprising to me that cancer was not positive and statistically significant; however, this is likely a result of the variety of cancer diagnoses in terms of severity, population affected, and available treatments. Still, there was a positive correlation to investment, but this was not statistically significant. Neurological treatments were observed to be statistically significant in a negative manner, with participation in this field being associated with less investment. I believe that this is due to the competitive nature of this subcategory of disease targets. Because Alzheimer's and other neurodegenerative diseases tend to be so widespread, there is a lot of research effort being put into finding a cure, or at least a way to mitigate symptoms, for it. As a result, many firms entered the race to address these diseases with gene editing when the technology became available, leading to increased competitiveness and lower chances of a given company "winning the arms race."

Results of this regression showed investments in blood and skin related treatments to be positively significant at the 10% level, while agriculturally based genetic editing companies tended to be negatively significant. The latter is likely the case simply because of the great inherent value in all of the other treatments addressing directly debilitating diseases that have been previously uncured in comparison to the value in agricultural products. Still, one benefit of an investment in agriculture is its quick turnover in terms of ideation to product stages, as while FDA approval is requisite, no direct clinical treatments are necessary (the product just has to be approved as safe for human consumption), which saves a significant portion of time. Interestingly, virological treatments showed great significance and were very positively correlated with more investment in a therapy. Therapies that were a part of this category were expected to receive over $300 million more. I see this as a result of the fact that the types of viruses requiring a gene editing based cure are likely extremely debilitating, transferable, and not curable through any other means. Also, CNS and musculoskeletal related treatments demonstrated a positively significant relationship with investment received.

Further analysis also revealed the round in which an investment occurred to be statistically significant. Seed rounds demonstrated the greatest and most consistent significantly between regressions, with the round associated with a $70 million decrease compared to the benchmark Series A. Compared to Series A, Series B rounds were significantly associated with a roughly $30 million dollar increase and Series C rounds were significantly associated with around a $64 million increase in investment.

Additionally, regression analyses revealed the relationship between investment year and investment amount. **Figure 8** demonstrates that collectively, there is no significant effect of the year of investment on the amount of investment a company receives. This was not what I anticipated, as I expected either macroeconomic variations to affect investment behaviors or hype in the industry to build over some course of time. In actuality, the coefficient associated with an investment occurring in 2020 is lower in comparison to other years, suggesting that macroeconomic conditions incited by COVID may have impacted investment that year.

Moreover, the coefficients associated with an investment occurring in 2016 were also lower relative to other years; this year, US GDP growth dropped to a low 1.7% (World Bank Data), suggesting that macroeconomic conditions may have some effect on investments.

Moreover, the prospect of hype seems to be somewhat present in these tables, as the post-COVID years 2021 and 2022 seemed to be correlated to an increase in investment amounts. These years represented a general boom in the biotechnology industry, as in response to the devastation of COVID and technological innovation, many investors sought out biotech startups. However, a lack of overall significance suggests that the gene editing sector is relatively unaffected by year in the grander scheme, and this agrees with the results from previously mentioned paper by Ledley. This is likely because biotechnological venture capital firms tend to have a unique life cycle in that investment times are premeditated and occur every 7 to 10 years, as opposed to the more commonly used 3 to 5 year cycle that exists in many other industries (Kang, 2018). Thus, it seems that when it is time for a venture investment group to begin its investment cycle, it will find ideal options that are presently seeking investment.

Figure 9: Investment (million USD) Regressed with Sentiment Buckets

	Positivity	Neutrality	Negativity	Emotion
VARIABLES	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)
Language Bucket 2	-21.20	-20.08	-16.00	4.860
	(24.49)	(22.31)	(27.21)	(29.83)
Language Bucket 3	-39.98*	4.063	17.97	-10.11
	(21.30)	(21.53)	(32.69)	(28.86)
Language Bucket 4	-11.12	-0.632	58.39	3.242
	(27.23)	(27.65)	(44.13)	(28.42)
PhD	98.26*	76.99	86.53	87.06
	(54.97)	(53.51)	(52.76)	(53.18)
Top 20 School	-162.3*	-135.4	-132.3	-165.4*
	(95.37)	(96.32)	(97.09)	(96.75)
Prior Company Found	366.7**	291.9**	294.3**	340.9**
	(145.8)	(146.4)	(144.8)	(148.0)

Prior Funding (mil USD)	0.104	0.122	0.110	0.103
	(0.173)	(0.174)	(0.177)	(0.175)
	Positivity	Neutrality	Negativity	Emotion
TARGET EFFECT	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)
Drug Delivery	-186.3*	-139.2	-132.3	-183.4*
	(102.1)	(106.0)	(113.6)	(102.0)
Cancer	252.3	201.9	185.7	229.9
	(158.1)	(155.6)	(156.4)	(160.0)
Rare Genetic Disease	106.2	58.11	74.94	59.18
	(81.99)	(82.67)	(82.64)	(83.50)
Ophthalmological	210.5	159.9	149.8	186.4
	(137.5)	(136.8)	(138.3)	(139.5)
Neurological	-182.0**	-142.1*	-147.0**	-156.9**
	(75.37)	(78.50)	(73.45)	(79.22)
Blood/Skin	254.7*	222.3*	182.8	232.6*
	(133.3)	(135.0)	(134.5)	(136.4)
Agricultural	-257.5***	-217.8**	-209.2**	-257.1**
	(94.50)	(99.77)	(94.98)	(101.7)
Virological	353.0***	308.9***	311.5***	317.1***
	(94.72)	(92.32)	(90.11)	(93.55)
CNS	256.2**	186.7*	201.7**	215.5**
	(99.97)	(98.36)	(95.36)	(98.78)
Musculoskeletal	300.0**	227.8	241.1	265.3*
	(149.8)	(152.3)	(148.9)	(150.4)
Organs	293.8	227.7	237.3	270.1
	(185.0)	(180.7)	(180.4)	(181.3)
Miscellaneous	-95.70	-92.12	-79.23	-111.3
	(73.40)	(76.29)	(78.77)	(77.25)
	Positivity	Neutrality	Negativity	Emotion
ROUND EFFECT	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)
Seed Round	-69.43***	-69.60***	-69.25***	-71.54***
	(17.41)	(17.58)	(17.47)	(17.68)
Series B	32.43*	29.75*	34.68**	32.60*
	(16.80)	(17.07)	(17.31)	(17.14)
Series C	67.26**	61.63*	69.31**	66.93**
	(31.19)	(31.64)	(32.33)	(31.93)
	Positivity	Neutrality	Negativity	Emotion
YEAR EFFECT	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)	Investment (mil USD)
2014	1.730	-5.939	-6.029	9.602
	(41.28)	(41.56)	(41.23)	(51.89)
2015	-4.068	-9.187	-7.360	-2.617
	(32.45)	(33.76)	(32.48)	(41.93)
2016	-35.31	-40.61	-40.73	-6.105
	(29.62)	(29.94)	(30.35)	(33.63)
2017	-12.56	-18.15	-15.36	-36.66
	(25.88)	(26.20)	(25.90)	(30.42)
2018	-20.95	-24.54	-24.52	-14.96

	(21.83)	(22.09)	(22.35)	(26.37)
2020	-23.71	-28.32	-26.99	-23.18
	(22.51)	(22.90)	(23.76)	(22.17)
2021	25.83	22.68	24.69	-26.73
	(24.38)	(24.64)	(25.37)	(22.93)
2022	33.37	27.23	25.08	23.32
	(33.93)	(34.22)	(34.23)	(24.67)
2023	-2.449	-0.965	-2.946	27.39
	(32.09)	(32.05)	(32.09)	(34.14)
2024	-61.62	-51.40	-54.99	-4.082
	(92.98)	(91.81)	(88.47)	(32.56)
Constant	-101.8	-70.59	-85.33	-86.98
	(87.12)	(88.68)	(84.93)	(86.05)
Adjusted R^2	.8387	.8353	.8375	.8341
Observations	277	277	277	277
Number of Round	4	4	4	4

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The introduction of language buckets demonstrated somewhat similar trends to those observed in **Figure 8**. For instance, all buckets of positivity had negative coefficients, suggesting that compared to the benchmark bucket encapsulating lower positivity, using more positive language was commonly associated with less investment received. Also, statistical significance was observed for Bucket 3 of positivity, suggesting that firms that use a substantial amount of positive language. Specifically, a firm's text having a 50 to 74.99% positivity score was observed to result in a decrease in investment by 39.98 million USD. Interestingly, while bucket 4's measurements were not significant, a firm's text's participation in it was markedly lower in magnitude at -11.12 million USD, suggesting that the extent to which positivity is used can affect the investment amount received. Outside of this bucket, there was no other significance in language buckets.

The negativity buckets showed an interesting trend. As the level of negativity increased, the correlation to investment also increased. While negativity bucket 2 was associated with a 16 million USD loss in investment, buckets 3 and 4 were shown to be associated with 17.97 and

58.39 million USD increases in investment. This evidence seems to substantiate earlier data from **Figure 8**, suggesting that there may be some sort of increased appeal by companies that utilize more negative language. Still, since there is no significance, it is fair to say that companies are not being swayed by the use of negative wording. Emotional and neutral buckets demonstrated little consistency between buckets. It is plausible that this occurred for these buckets because of the differential impacts of other positive or negative language, which have more established trends, in the analyzed text. Additionally, the inconsistencies associated with these buckets, in addition to its lack of statistical significance, indicate that investors are largely not being affected by neutral or emotional language.

Furthermore, the fact that there is no statistical significance of any bucket in negativity, neutrality, or emotion, suggests that there no incentive for a firm to use language that is in one bucket of these sentiments. This further substantiates the lack of significance observed in **Figure 8**. Interestingly, **Figure 9** having statistical significance for bucket 3 of positivity support the trends lacking significance observed in **Figure 8** that it might be beneficial for firms to avoid positive language. **Figure 9** specifies this, demonstrating it is not ideal for firms to use wording that is somewhat positive, perhaps because this type of language can be interpreted misleading and unrealistic. I postulate that the same significance cannot be established for the more positive bucket because some firms within this bucket could have more legitimately reasons to be positive and do not appear as if they are trying to persuade investors with their wording.

Unsurprisingly, many of the other variables remained relatively consistent between both regressions. In terms of founder value, some changes in significance were observed, but general implications remained the same. In **Figure 8** in which positivity and negativity demonstrated positive statistical significance, whereas in **Figure 9**, only negativity did so. Moreover, while in

Figure 8, a significant negative correlation was observed between attendance of a Top 20 school and investment in the negativity regression, this was observed in only the positivity and emotion-based regressions in **Figure 9**. These variations largely took place due to slight differences in p-values of these variables across regressions bringing certain coefficients marginally closer to or away from the 10% level of significance. In general, the two figures in combination seem to substantiate previously the established trends of positive value associated with PhDs and negative value associated with Top 20 school attendance. Additionally, the value of having founded a prior company maintained similar magnitude and significance throughout both regressions.

Outside of these variables, there were many consistencies observed between the effects for disease, round, and year between the two regression models. This is a positive sign that between these fixed effect models, the variables are likely encompassing the appropriate change in investment associated with them. The values and significance of all of the prior funding, year effect, target therapy effect, round effect, and firm effect are very similar between these models. Also, the constants generally remain around the same value, likely for the same reasons previously mentioned. R^2 values show very minimal improvement, suggesting that the categorization of words slightly improves the predictability of this model. This is sensible due to similarities in data of the two regressions. Because of this, I do not find one form of analysis preferential to the other.

IV. Conclusion

In conclusion, I created a dataset consisting of data points representing venture investments in gene editing startups. This dataset consisted of various firm-based characteristics such as founder traits, therapeutic targets, investment year, investment round, firm ID, and prior funding. Most importantly, though, to assess how verbiage a company uses might affect its investment, I pulled excerpts from company websites' home pages prior to the time of a given investment and put them through an AI NLP based sentiment analyzer. While this tool's lack of explanation of its methodology was subject to some concern, I aimed to address these concerns by understanding and subsequently optimizing my usage of the tool to procure data as accurately as possible. I then ran regressions for AI values given relating to positivity, negativity, neutrality, emotion, and adjective count. Additionally, I ran a regression in which I categorized positivity, negativity, neutrality, and overall emotion in order to group texts by generic sentiment. Both regressions largely suggest that investors were not swayed by the language that companies utilize. However, the categorical regression with buckets of language did imply negative significance associated with somewhat positive verbiage (50% to 74.99% positivity. Firms using this type of language were observed to be correlated to a decrease in investment received.

While these results were not significant, regression analyses revealed that having a PhD and having founded a prior company is significantly associated with more investment, while having attended a top institution was significant in some regressions in a manner than it was associated with less investment. Moreover, certain target diseases were significantly associated with changes in investment behaviors, suggesting that value can be at least partially evaluated by the target of a therapy and quantifying the extent to which certain targets affect investor appeal.

Also, investment round was found to be associated to investment to a significant degree, as expected, while investment year was found not to be.

VI. References

Gostimskaya I. CRISPR-Cas9: A History of Its Discovery and Ethical Considerations of Its Use in Genome Editing. Biochemistry (Mosc). 2022 Aug;87(8):777-788. doi: 10.1134/S0006297922080090. PMID: 36171658; PMCID: PMC9377665.
"Genome Editing Market Size, share & Trends Analysis Report By technology (CRISPR, ZFN, TALEN), by delivery method (Ex-Vivo, In-Vivo), by application, by mode, by end-use, by region, and segment Forecasts, 2023 - 2030." 2023. May 15, 2023. https://www.grandviewresearch.com/industry-analysis/genome-editing-market#:~:text=Genome%20Editing%20Market%20Size%20%26%20Trends,17.8%25%20from%202023%20to%202030.

FDA Approves Cure for Sickle Cell Disease, the First Treatment to Use CRISPR." 2023. NBC News. December 8, 2023. https://www.nbcnews.com/health/health-news/fda-approves-cure-sickle-cell-disease-first-treatment-use-gene-editing-rcna127979.

Sheetz, Michael. 2018. "Secretary DeVos, Walmart Heirs and Other Investors Reportedly Lost Over $600 Million on Theranos." CNBC, May 4, 2018. https://www.cnbc.com/2018/05/04/theranos-devos-other-investors-reportedly-lost-over-600-million.html.

Collins, Karen, Hannah Johnston, Adel Manji, and Bill Kapralos. 2023. "Presentation Mode and its Impact on Sentiment in Free Verbalization Responses to Sounds." *SSRN*, January. https://doi.org/10.2139/ssrn.4670724.

World Bank Open Data. (n.d.). World Bank Open Data. https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=US

Kang, Hyunsung D. 2018. "A Start-Up's R&Amp;D Stages and the Evolution of Financing Sources: Evidence From the Biotechnology Industry." *Entrepreneurship Research Journal* 8 (3). https://doi.org/10.1515/erj-2017-0159.

Weinstein, A., Sipala, A., Turkington, L., & Stromberg, M. (2016). Theranos - A case study on customer value and technology. *Journal of Marketing Perspectives, 1*, 6-22.

Office of the Commissioner. (2023, December 8). *FDA Approves First Gene Therapies to Treat Patients with Sickle Cell Disease*. U.S. Food And Drug Administration. https://www.fda.gov/news-events/press-announcements/fda-approves-first-gene-therapies-treat-patients-sickle-cell-disease

Godfrey, P. C., Allen, G. N., & Benson, D. (2020). The biotech living and the walking dead. *Nature Biotechnology*, *38*(2), 132–141. https://doi.org/10.1038/s41587-019-0399-1

Evens, R. P. (2023). Biotechnology Commitment by Pharma Industry and its evolution vs Biotech Companies in Healthcare, Drug Development, and Business, Past & Present [2001-2022]. *Medical Research Archives*, *11*(5). https://doi.org/10.18103/mra.v11i5.3876

Fletcher, L. (2023, June 3). *Unlocking the future: where is gene editing going next? - Front Line Genomics*. Front Line Genomics. https://frontlinegenomics.com/unlocking-the-future-where-is-gene-editing-going-next/

Ledley, F. D., McNamee, L. M., Uzdil, V., & Morgan, I. (2013). Why commercialization of gene therapy stalled; examining the life cycles of gene therapy technologies. *Gene Therapy*, *21*(2), 188–194. https://doi.org/10.1038/gt.2013.72

Ke, Q., & Wang, W. (2013). The adjective frequency in advertising English slogans. *Theory and Practice in Language Studies*, *3*(2). https://doi.org/10.4304/tpls.3.2.275-284

Hanna, E. Y., Rémuzat, C., Auquier, P., & Toumi, M. (2017). Gene therapies development: slow progress and promising prospect. *Journal of Market Access & Health Policy*, *5*(1), 1265293. https://doi.org/10.1080/20016689.2017.1265293

Pian, J., Chandra, A., & Stern, A. D. (2020). The past, present, and (Near) future of gene therapy and gene editing. *NEJM Catalyst Innovations in Care Delivery*, *1*(5). https://doi.org/10.1056/cat.20.0072

Ranjan A., Jian N. (2015). Avoiding failures in biotechnology start-up. https://research-api.cbs.dk/ws/portalfiles/portal/58510840/ashish_ranjan_og_nanzi_jiang.pdf

Godfrey, P. C., Allen, G. N., & Benson, D. (2020b). The biotech living and the walking dead. *Nature Biotechnology*, *38*(2), 132–141. https://doi.org/10.1038/s41587-019-0399-1

Baum, J. a. C., & Silverman, B. S. (2004). Picking winners or building them? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups. *Journal of Business Venturing*, *19*(3), 411–436. https://doi.org/10.1016/s0883-9026(03)00038-7

Block, J., Diegel, W., & Fisch, C. (2023). How venture capital funding changes an entrepreneur's digital identity: more self-confidence and professionalism but less authenticity! *Review of Managerial Science*. https://doi.org/10.1007/s11846-023-00686-1

Zaruba, N., Petrova, M., Vostrikov, K., Wolfson, E., & Nikiforova, O. (2021). Actual strategies of interaction between government bodies and mining enterprises: personnel problems. *E3S Web of Conferences*, *315*, 04002. https://doi.org/10.1051/e3sconf/202131504002

M. Siering, ""Boom" or "Ruin"--Does It Make a Difference? Using Text Mining and Sentiment Analysis to Support Intraday Investment Decisions," *2012 45th Hawaii International Conference on System Sciences*, Maui, HI, USA, 2012, pp. 1050-1059, doi: 10.1109/HICSS.2012.2.

Bi, S., Liu, Z., & Usman, K. (2017). The influence of online information on investing decisions of reward-based crowdfunding. *Journal of Business Research*, *71*, 10–18. https://doi.org/10.1016/j.jbusres.2016.10.001

McLeod, M. S., Sears, J. B., Chandler, G. N., Payne, G. T., & Brigham, K. H. (2022). Rhetoric, risk, and investment: letting the numbers speak for themselves. *Journal of Management Studies*, *59*(7), 1657–1687. https://doi.org/10.1111/joms.12812

Palada, B. (2020). Bright outlook for the global cell and gene therapy market. *Genetic Engineering & Biotechnology News*, *40*(S2), S6–S8. https://doi.org/10.1089/gen.40.s2.02

Carr, D., & Bradshaw, S. (2016). Gene therapies: the challenge of super-high-cost treatments and how to pay for them. *Regenerative Medicine*, *11*(4), 381–393. https://doi.org/10.2217/rme-2016-0010

Hand, J. R. M. (2007). Determinants of the round-to-round returns to pre-IPO venture capital investments in U.S. biotechnology companies. *Journal of Business Venturing*, *22*(1), 1–28. https://doi.org/10.1016/j.jbusvent.2005.09.004

Taffler, R., Spence, C., & Eshraghi, A. (2017). Emotional economic man: Calculation and anxiety in fund management. *Accounting, Organizations and Society*, *61*, 53–67. https://doi.org/10.1016/j.aos.2017.07.003

Kristensen, U. (2023, March 9). *The financial potential of gene editing therapy: Breaking down the $18B investment*. Emersion Insights. https://emersioninsights.io/the-financial-potential-of-gene-editing-therapy-breaking-down-the-18b-investment/

VIII. Appendix

Writing Sample #1: Theranos (Not a Data Point) - 2016

"The lab test, reinvented. We believe the future of health care lies in greater access for the individual. So we built a better lab experience with access in mind, making it easier than ever for you to engage with your health early and at the time it matters most.

Experience the Theranos difference. We've made getting a lab test fast, convenient, affordable — and a lot more human. Two simple steps. Get a lab order. Ask your physician to write you a lab order, or download one of our forms and bring it to your next visit. Get tested. Bring any lab order to a Theranos Wellness Center™ located inside a participating Walgreens. We're open early, late, and on the weekends.

The same low prices for everyone. Whether you have good insurance, bad insurance or no insurance at all, you should be able to afford lab testing. Which is why we charge everyone the same low prices. Period. Prices that are always a fraction of the cost of all other labs. And we publish them right here on our site. Meaning there are no surprises, and you know exactly what you are paying, before you get tested.

Smaller samples. Smaller needles. A better experience. Our tests, including venous draws, require smaller samples than traditional labs. We also use much smaller needles. Ones designed specifically for collecting blood for children, taking the smallest sample possible. Theranos tests mean less blood, an easier process, and a clear difference in your experience.

Your health, in your hands. They're your results. You should have access to them. With the Theranos app you and your physician receive private, highly secure results faster than ever before, to view any time day or night. Since we've made it easier to get access to the tests you need, you and your physician can better track small changes that emerge over time. To help provide insight into where your health is headed, and the early detection of disease."

Positivity	Negativity	Neutrality
23.53 %	11.76 %	64.71 %

Writing Sample #2: Fractyl Health - 2014

“PIONEERING BREAKTHROUGH THERAPIES TO TREAT TYPE 2 DIABETES (T2D) AND OBESITY

Our vision is to develop transformative therapies that have the potential to prevent and eliminate metabolic diseases. We are a mission-driven team of innovators, singularly focused on developing therapies that are designed to target root causes of T2D and obesity and delivering those therapies to patients as broadly and rapidly as possible.

Revita is an outpatient procedural therapy designed to durably modify duodenal dysfunction via hydrothermal ablation in order to restore metabolic health.

Rejuva is a novel, locally administered, AAV gene therapy platform currently in preclinical development to improve islet function. We believe the platform represents a major advance in incretin-based therapies. PGTx is designed to potentially enable long-term remission of T2D and obesity by durably altering metabolic hormone response in the pancreatic islet cells of patients with those diseases.

Led by world-renowned experts in metabolic disease, the EraseT2D task force shares Fractyl Health’s mission to build evidence, understand, and ultimately eradicate metabolic disease.

The Erase T2D task force builds on decades of discovery that have clearly identified the critical role of the gut as a regulator of metabolic disease. The signaling mechanisms between the gut and the rest of the body are numerous and not yet fully understood. The task force is charged with advancing research on the role of the gut in metabolic disease to catalyze future discoveries that may inform how diabetes can be better understood and ultimately erased.
This initiative, funded by Fractyl Health, is meant to spur the kind of innovative thinking and fundamental breakthroughs that will point the way toward a world without type 2 diabetes.”

Positivity	Negativity	Neutrality
66.67 %	8.33 %	25 %

Writing Sample #3: Verve Therapeutics - 2020

“Protecting the World from Heart Disease. We are developing therapies to safely edit the adult genome and confer lifelong protection from cardiovascular disease.

Verve is focused on discovering and developing therapies that safely edit the genomes of adults to confer protection against coronary artery disease, the most common type of heart disease and the leading cause of death worldwide.

Verve brings together two of the biggest breakthroughs in 21st century biomedicine — human genetic analysis and gene editing — to realize a new future, one of longevity and vitality for tens of millions of people worldwide at risk of coronary artery disease.

Verve has been purpose-built for the task ahead, with a founding team of world-leading experts in cardiovascular medicine, human genetics, gene editing delivery technology and safety, and drug development.”

Positivity	Negativity	Neutrality
100%	0%	0%

Written Sample #4: Tango Therapeutics - 2017

“We are going beyond current cancer treatments and targeting unaddressed disease mechanisms to develop transformational new drugs for patients. Powered by genetic data from patient tumors and advances in CRISPR-based target discovery, we are discovering medicines that embrace the complexity of cancer with the bold goal of cures.

Accelerating discovery by putting patient selection first

We use deep DNA sequencing and CRISPR-based screening to identify targets for specific subgroups of cancer patients. Our goal is to leverage the principles of synthetic lethality to find the weaknesses in cancer cells created by genetic complexity and use them to provide a roadmap to cures.

Our discovery efforts begin and end with patients. We invert the traditional discovery paradigm by putting patient selection before target identification to provide the right medicines to the right patient to improve speed and success in drug development. Essential to this effort is depth of understanding of the genetic subtypes of cancer and the drug combinations uniquely relevant to each subtype. As we advance into the clinic, our trials will enroll the patients most likely to benefit from our new treatments, enabling speed, success and impact for patients.”

Positivity	Negativity	Neutrality
22.22 %	55.56 %	22.22 %