Industry-Funded Research and Bias in Food Science

Anita Rao*

Booth School of Business

University of Chicago

May 7 2021

First Draft: June 20, 2020

Abstract

Is industry-funded scientific research likely to be biased towards finding positive results? Is industry more likely to work on topics with likely positive outcomes? Using publication-level data and focusing on food groups that are typically considered healthy, I evaluate each article's abstract using crowdsourcing tools. I find little evidence to support selection on topics with positive outcomes, but industry is less likely to work on topics classified as unrelated to health. Conditional on a topic, I find that industry-funded research is 4.1% more positive compared to non-industry funded research with grains that receive heavier funding responsible for most of the effect. Industry-funded research is also more likely to receive a mention in certain industry newsletters. Coupled with firm incentives to use science to further their marketing efforts, such increased trade press coverage might play a role in shaping consumers' opinions on what is healthy.

KEYWORDS: empirical industrial organization; consumer protection; industry funding; bias in science

JEL CODES: L20, M31, O31

*Thanks to participants at the 2020 QME Conference, 2020 Marketing Science Conference and to seminar participants at the FTC Bureau of Economics, Wharton, UCSD and UIUC. Thanks to Richard Liu and Davit Musaleyan for excellent research assistance. Thanks to J.P. Dube, Guenter Hitsch and Ginger Jin for thoughtful comments. All errors are my own. Email: anita.rao@chicagobooth.edu

1 Introduction

Firms have incentives to conduct scientific research to convince consumers that their products are healthy. Firms also have incentives to convince policymakers and nutrition professionals - who evaluate various health claims in the marketplace and thereby influence consumer opinion - that their claims are accurate. These incentives can create two sources of bias. First, firms might conduct research on only certain topics with known positive outcomes. Second, conditional on a topic, firms might report/find favorable outcomes. A limited body of research has documented the second kind of bias, but has either used a case-study or meta-analyses based approach and has focused only on unhealthy products such as sugar-sweetened beverages (e.g., Vartanian et al. 2007 and Bes-Rastrollo et al. 2013). The first kind of bias has been reported only anecdotally (Nestle 2020). Moreover, a case-study based approach can only comment on the impact of industry-funding within a given topic and is not well-equipped to answer the question whether firms are conducting research on selective topics.

This paper aims to document the kind and existence of bias, if any, across a wide variety of food groups and research topics. Moreover, this paper focuses on food groups typically considered healthy: whole grains, to understand if the documented bias in unhealthy food products extends to healthy food groups as well. Further, to understand if such articles are likely to have shaped consumer opinion, this paper measures the likelihood of industry-funded articles reaching media outlets.

Using a comprehensive database of published scientific articles and restricting attention to *Food, Science & Technology*, and *Nutrition Dietetics*, categories that pertain to health outcomes, I extract detailed data on research articles published in all types of whole grains. This database provides the title, abstract, number of citations, and most importantly, the funding source for each research article. I then use crowdsourcing to classify each abstract as positive, negative, neutral, or unrelated to health outcomes. Finally, using Factiva- a news archive database—I collect media mentions per research article.

Using keywords to define a topic, I find little evidence supporting strategic selection of topics by firms. Industry-funded research is no more likely than non-industry funded research to conduct research on positive vs. negative research topics. However, they are 2% less likely to work on topics classified as unrelated to health outcomes. Examining the bias conditional on a topic, I find that industry-funded research is 4.1% more positive than non-industry-funded research. Moreover, there is substantial heterogeneity across grains, with the effect being largely driven by grains with heavy industry funding (corn, rice and wheat) where industry-funded research is 8%-9% more positive.

To establish a more causal relationship, I examine author-topic pairs where the same author for a given topic receives industry-funding for one research article and no industry-funding for the other. The main result continues to hold in this stricter subset of the data.

To understand whether the positive bias is driven by false and unsubstantiated claims, I examine the quality of research articles. I find little evidence to suggest that industry-funded research is of lower quality. First, industry-funded articles receive 1.36 more citations than non-industry funded articles and are also published in journals with slightly higher impact factors. Second, a textual analysis of the abstracts shows that industry-funded research is more likely to use words such as *randomized* and *experiment* indicative that the research design being used is randomization, a gold-standard in the scientific literature.

I then examine whether industry-funded articles are more application-oriented and/or less scientifically written. I find no evidence that industry-funded abstracts use fewer scientific words. However, industry-funded abstracts are on average longer, containing 14 more words. One potential explanation for the longer text is that industry-funded articles use sentences that provide an interpretation of the findings towards applicability to human health outcomes.

Examining media coverage, I find that industry-funded articles are more likely to garner mentions in certain industry newsletters (e.g. *Health & Medicine Week*). Combined with the findings that industry selects on topics that are more related to health outcomes and that industry-funded articles have longer abstracts indicative of an application-oriented focus, this increased trade press coverage suggests incentives of industry to disseminate their work and to conduct scientific research as an avenue of influencing policy makers and nutrition professionals who likely play a role in shaping consumer opinion.

Contribution

Nutrition research and industry sponsorship

Industry-funded research and the extent of its bias on outcomes has received recent empirical attention in the nutrition sciences. Almost all of this research is directed toward documenting how industry-funded research finds less harmful effects of its products. Sugar, sugary products, and sugar-sweetened beverages have received most of the attention (Lesser et al. 2007, Massougbodji 2013, Kearns et al. 2016) followed by olestra – a highly controversial fat substitute - and GMOs (Levine et al. 2003, Diels et al. 2010).

This area is still relatively unexplored empirically. Nestle (2018) cites finding only 11 scientific studies examining the relation between industry-funding and food and beverage research through 2018. In contrast, there are thousands of studies on how the pharmaceutical industry influenced scientific research (Belluz 2018).

The extant research does not document the relationship between products typically considered *healthy* and industry-funding. Two exceptions are Nkansah et al. (2009), who look at calcium supplementation, and Wilde et al. (2012), who look at dairy consumption. However, small-sample-size issues or the fact that industry was almost always involved made identification hard in these two studies. Using a large database of research articles in the specific categories of *Nutrition Dietetics* and *Food, Science & Technology*, this paper contributes to the extant literature by looking at products that consumers typically consider healthy.

Analyzing data across multiple food groups and research topics, rather than take a case-study approach as is done in the extant literature, allows me to additionally answer the question whether industry is likely to conduct research on selective topics, e.g. topics with likely positive outcomes. A case-study approach, on the other hand, can only comment on the impact of industry-funding within a given topic.

This paper also analyzes differences in various quality metrics using journal of publication and citation data. The extant literature implies industry-funded research is of lower quality of research. In contrast, Myers et al. (2011) find little evidence that industry-funded research is of lower quality. The nature of this relationship in healthy products is less clear. In unhealthy products, firms have incentives to refute research that shows their products in an unfavorable light. In healthy products, if extant non-industry funded nutritional research is already positive, firms might have fewer incentives to make unsubstantiated claims.

Furthermore, this paper examines the propensity of such research reaching research-centric media outlets. Presence in such media outlets is a proxy for firm-driven public relations (PR) efforts and is suggestive of firm efforts to influence consumer opinion.

Pharmaceutical research and industry sponsorship

Research showing the link between industry sponsorship and pharmaceutical research predates that of the nutrition sciences. Three systematic reviews in the medical literature (Bekelman et al. 2003, Lexchin et al. 2003, Sismondo 2008a) find pharmaceutical industry funding is associated with favorable outcomes. Explanatory factors suggested in Sismondo (2008b) and Lexchin et al. (2003) include publication bias (under-reporting of unfavorable outcomes) and design bias (use of poor comparators, poor end-points or selected trial duration that would not show side effects). More recently, Oostrom (2021) shows that for a given drug and trial, industry-funded trials report more positive outcomes. Unlike pharmaceutical research, which is meant to be consumed and evaluated by experts of the field (i.e., doctors), nutrition research that reaches consumers (either through media or marketing efforts) is directed at a more susceptible population. Moreover, while pharmaceutical products face considerable scrutiny by regulatory authorities before entering the market, consumer product goods exist in a lesser regulated space.

Selective information disclosure

This paper is broadly related to the work on selective information disclosure. Moorman (1998), Burke et al. (1997) and Hastak and Mazis (2011) document various types of truthful but misleading deceptive practices used by firms (e.g. "contains oat bran" might imply a substantial amount of oat bran; "no cholesterol" might imply competitors contain cholesterol). This paper contributes to this literature by documenting firm-driven research geared towards (perhaps selectively) showcasing its products' positive attributes.

That industry ownership and sponsorship can bias incentives has been documented in other domains as well (e.g., Dellavigna and Hermle 2017, Reuter and Zitzewitz 2006). Dranove and Jin (2010) provide a review of quality disclosure and seller incentives to voluntarily disclose information.

2 Data and Descriptives

In this section, I explain the three main sources of data used to answer the research question: the database of research articles containing the abstract and funding information, the crowdsourcing procedure used, and the news archive database used to extract news mentions of each research article.

2.1 Research article abstracts and attributes: Web of Science

Research article abstracts are obtained from Web of Science, a database of peer-reviewed scientific articles across various disciplines. Each article's title, abstract, journal and year of publication, citation count, and funding source are obtained from Web of Science.

Most journals began requiring funding disclosures only recently. Krimsky and Rothenberg (2001) state that only 16% of highly ranked scientific and biomedical journals had conflicts-of-interest policies in place during 1997. Nearly a decade later, a survey conducted by Cooper et al. (2006) on biomedical journals showed that 93% of journals had such policies in place. This pattern is observed in the data used in this paper as well, with 99% of disclosures occurring post 2007. Almost all industry-funding disclosures are observed post 2007. I therefore restrict attention to data post 2007.

Restricting attention to whole grains, which are unequivocally considered healthy, I identify 22 whole grain groups, as listed by the Oldways Whole Grains Council[1], ranging from amaranth to wild rice. For each of the 22 grains, I identify alternative names to be comprehensive in the search for articles; for example, corn is also known as maize and oats as oatmeal.

I then conduct a search on the Web of Science database, using the names of the identified food groups and restricting attention to the categories of *Nutrition Dietetics* and *Food Science & Technology*. These two categories pertain most to health and nutrition. Research on whole grains is conducted in other categories such as *Plant Sciences*, *Agronomy*, *Agriculture*, and *Soil Sciences*, but these categories are irrelevant for our purposes because they do not pertain to health. The top journals, where nearly 50% of research articles are published in this database, are *Food Chemistry*, *Journal of Cereal Science*, *Journal of Agriculture and Food Chemistry*, *Cereal Chemistry*, *Journal of the Science of Food and Agriculture* and *LWT-Food Science and Technology*.

Each funding source is classified as *industry* or *not industry* based on the information contained in the funding text. Three main sources of data help in such classification: the Top 100 Food and Beverage Companies from Food Processing, a list of companies from ReferenceUSA and Food Manufacture Directory, and a comprehensive list of industry suffixes across countries, for example, inc, co, ltd, spa (Italy) and gmbh (Germany). Industry-funded research forms a small portion of all research, with the median across 21 grain food groups (the whole grain freekeh had no research articles) having 1.6% industry funded articles.

Table 1 presents the total number of abstracts and the percentage of industry-funded articles in each whole grain. For clarity of presentation in the table, grains with fewer than 100 abstracts are excluded (wild rice, bulgur, einkorn, farro, fonio, kaniwa, khorasan wheat, spelt and teff). However, they will be included in the analysis. Unsurprisingly, staple foods such as rice, wheat, and corn receive a lot of academic attention, evidenced by the absolute number of research articles in these grains. These grains also have more industry-funded articles. Oatmeal receives the highest percentage of industry funding with 13% of all abstracts being industry-funded. Table 1 also presents the top industry funder and the number of research articles it funds in that grain category.

[1]https://wholegrainscouncil.org/whole-grains-101/whole-grains-z Accessed April 29, 2020

Table 1: Percentage Industry Funding in Whole Grains Food Sciences

grain	Total Abstracts	% industry	Top industry funder (associated abstracts)
oats or oatmeal	860	13.1%	pepsi (37)
rye	355	11.8%	lantmannen (9)
barley	1,320	11.0%	american malting barley association (22)
corn	4,121	5.1%	cargill (13)
wheat	5,624	5.0%	lantmannen (18)
sorghum	757	4.6%	brazilian agricultural research corp. (3)
triticale	179	3.9%	grains research & development corp. (2)
rice	5,394	2.8%	dsm (5)
millet	504	1.6%	belisle solution nutrition (2)
buckwheat	539	1.3%	bioglane (1)
amaranth	322	1.2%	nestle (2)
quinoa	330	1.2%	andean naturals (1)

2.2 Abstract classification: Crowdsourcing

The next step involves classifying each research abstract as positive, negative, or neutral as it relates to health outcomes. Unlike movie or product reviews, which have been extensively classified using machine-learning tools such as sentiment analysis, scientific abstracts carry little to no sentiment: even positive findings are couched in neutral tones and words such as positive and significant might not translate into a positive finding as it relates to health outcomes. Figure 1 depicts the first and last sentences of an abstract, the classification of which requires knowledge that (a) cholesterol is undesirable and (b) lower cholesterol is better for health. Other examples of sentences containing relevant health-outcome information and their correct classification are presented in Table 2.

On the other extreme, case studies and systematic reviews that use specific outcomes such as energy intake and body weight, are severely limited in the number of articles they can study. For example, Vartanian et al. (2007) examine 88 articles and Bes-Rastrollo et al. (2013) examine 17 systematic reviews. Because one of the goals of this paper is to examine the impact of funding across a wide variety of articles and food groups, examining a particular type of health outcome is too restrictive for the purposes of this paper.

I therefore resort to measuring health outcome information contained in abstracts and use crowdsourcing tools to classify various research abstracts. Amazon's Mechanical Turk provides an efficient way to crowdsource classification. MTurk workers - Amazon's term for those who sign up to work on various Human Intelligence Tasks (HITs) - were asked to classify abstracts as positive, negative, neutral, or unrelated as related to a health outcome. Workers were recruited for sentiment analysis HITs with the title "Scientific Journal Abstract

Analysis (~2 minutes to read and analyze)" and were remunerated \$0.15 per task with an additional \$0.15 bonus awarded for accuracy. Five independent workers were assigned to each task to get a higher degree of accuracy.

Table 3 provides the average rating across abstracts by funding source. Abstracts are coded as 1 for positive ratings, 0 for neutral ratings, and -1 for negative ratings. Unrelated ratings, unless otherwise specified, are treated separately. The crowdsourced ratings data is collected in two phases. Because oatmeal has the highest percentage of industry funding, I begin with a case study of oatmeal to test, in a descriptive manner, whether industry-funded research is associated with higher ratings. Online Appendix A presents the results of this descriptive analysis, after which the second phase was launched. In the second phase, using stratification, I select a sample of industry- versus non-industry-funded articles to obtain crowdsourced data. Use of this statistical sampling procedure is necessary because crowdsourcing the nearly 20,000 abstracts across all remaining whole grains is cost prohibitive. I therefore resort to sampling a smaller subset of articles per food group. Complete random sampling is sub-optimal in this setting because industry-funded articles form a small portion of all articles per food group, with the median at 1.6%. Therefore, a random-sampling procedure that ignores this information will have very few articles that belong to the industry strata. Instead, a procedure known as stratification, where the population is divided into sub-populations and each sub-population is sampled separately, is more efficient (Piazza 2010). Piazza (2010) points out that the two main reasons for stratification are facilitating estimation within a sub-group of interest and increasing sample precision. Using stratification - more specifically, disproportionate sampling - I sample at a 100% rate from industry-funded articles and a lower 10% rate from non-industry-funded articles. Doing so results in 2,734 abstracts, which are then crowdsourced for classification as positive/negative/neutral[2]. All analysis that uses this stratified sample will be appropriately weighted.

The average rating across all abstracts is 0.39 in oatmeal and 0.14 in other whole grains, suggesting a skew towards positive findings. Unlike research on unhealthy substances such as sugar, the healthy food groups studied here have, on average, positive outcomes. The split between industry and non-industry is 0.47 and 0.38 in oatmeal and 0.20 and 0.13 for other grains, suggesting industry-funded articles are of a more positive nature.

2.3 News mentions: Factiva

To understand if industry-funded articles receive more media attention (e.g., companies might have more resources for PR efforts), I collected data from Factiva on each article's

[2]That disclosures became prominent only post 2007 was discovered after crowdsourcing. A large sample of 3,715 abstracts, including articles pre 2007, was included in the data collection task.

news mentions. To do so, each article's title was put into Factiva's search engine and the number of news mentions extracted. Table 3 provides the average news mention (coded as yes or no) across abstracts by funding source. For comparison, the citation-count statistics are also provided.

The most common news journal covering research articles in whole grains is *Food Weekly News*, forming 42% of all covered articles. The top six sources, presented in Table 4, account for 81% of the coverage. Moreover, some of the news sources specifically cover industry-driven research, such as *Biotech Week*, which describes itself as covering the latest news from the biotech and pharmaceutical industries. These news sources therefore serve as a metric of firm-driven marketing/PR efforts.

Mainstream media outlets such as *The New York Times* and *The Wall Street Journal* constitute a small portion (less than 0.01%) of media mentions for the research articles in the database. To ensure this is not driven by the fact that mainstream media articles often do not cite the research article titles they quote but instead use generic terms such as "study finds" referencing author names and/or journal names, I conduct a manual search. Specifically, I conduct another search for articles with the word "study" and the name of the whole grain, restricting attention to the *Health* and *Nutrition* categories of *The Wall Street Journal* and *The New York Times*. I then manually look for matches between mentioned authors and/or journals in each news article and the abstract database. Such an exercise resulted in no new media articles. I also conduct an automated search using each article's authors' names along with the article's food group, manually checking if the covered media article pertained to the research article. This exercise resulted in a few additional non-industry-funded articles (forming less than 0.2% of all media mentions).

Background: Consumption of 3 g oat beta-glucan/d is considered sufficient to **lower serum LDL cholesterol**, but some studies have shown no effect. [...] Objectives: [...]. Design: [...] Conclusions: [...] **an extruded breakfast cereal containing 3 g oat beta-glucan/d** with a high-MW (2,210,000 g/mol) or a medium-MW (530,000 g/mol) **lowered LDL cholesterol** similarly by approximate to 0.2 mmol/L (5%), but efficacy was reduced by 50% when MW was reduced to 210,000 g/mol. [...]

Figure 1: First and last sentences of research abstract

Table 2: Example Sentences from Various Abstracts and Their Classificaation

No.	relevant sentence in abstract	classification	industry
1	oat beta-glucan could **regulate the glucose metabolism**	positive	no
2	Fortified oat beverages may offer a convenient and effective mechanism to **improve the iron status of children**	positive	yes
3	Consumption of a whole-grain RTE oat cereal as part of a dietary program for weight loss had **favorable effects on fasting lipid levels** and waist circumference	positive	yes
4	These changes in molecular structure partially explain the **reduced digestibility** and viscoelasticity.	negative	no

Table 3: Summary Statistics for Dependent Variables: Industry vs. Non-industry

Dependent variable	All	Industry	Non-industry	No. Obs
Case Study: Oats				
Rating (Positive: +1, Neutral: 0, Negative: -1)	0.39	0.47	0.38	850
Citations	13.91	18.36	13.22	850
Factiva news mentions (Yes:1, No:0)	0.54	0.48	0.55	850
All Other whole grains				
Rating (Positive: +1, Neutral: 0, Negative: -1)	0.14	0.20	0.13	2,734
Citations	12.58	12.97	12.56	2,734
Factiva news mentions (Yes:1, No:0)	0.54	0.57	0.54	2,734

Note: Abstracts rated as "Unrelated" are assigned a numerical value of 0 for the ratings calculation in this table. Appropriate sampling weights applied for the "All Other Whole Grains" sample.

Table 4: Top News Journal Sources

News Journal	**Coverage**	**Areas covered**
Food Weekly News	42%	food science
Life Science Weekly	17%	animal and plant science
Agriculture Week	12%	farm news
Chemicals & Chemistry	4%	chemical industry
Biotech Week	3%	biotech and pharmaceutical industries
Health & Medicine Week	3%	industry of interest to health-minded professionals or consumers.
Covered research artciles	10,320	
Total research articles	19,089	
No. of news journals	201	

Note: This table uses the entire set of research articles (i.e., before sampling) in the Web of Science database and their corresponding news mentions.

3 Firm Incentives and Sources of Bias in Science

There are four main players in the market for food products: consumers, nutrition professionals, firms and regulators. Below I describe consumer preferences and regulators' priorities that can influence firm behavior.

Consumers All else equal, consumers prefer eating healthy foods. Consumers are persuaded by advertising, media and scientific articles, but trust the media and science more than they do advertising.

Nutrition Professionals Nutrition Professionals recommend healthy diets to individuals. They read peer-reviewed journals and research-centric newsletters to stay up-to-date on the latest research in the food sciences.

Regulators Regulators aim to protect consumers from fraudulent and deceptive practices in the marketplace. Regulators are on the lookout for false health claims and might look to peer-reviewed journals for evidence for or against such health claims.

Firms Firms have incentives to convey their products' healthy attributes. Typical means of such communication include advertising and front-of-package labeling. However, because consumers are more skeptical of such advertising, firms have incentives to convey the healthiness of their products through a trustworthy source. Engaging in scientific research is one way to increase credibility. Other means include citing scientific literature that supports the firm's health claims[3]. Firms also have incentives to conduct scientific research to persuade regulators and nutrition professionals.

Firms then disseminate this research through various marketing tools so that consumers as well as nutritional professionals can easily acquire this information. As an example, in its website, Quaker highlights its participation in scientific research through the Quaker Oats Center of Excellence[4]. The website consists of resources for consumers as well as nutrition professionals and provides links to various published articles which were funded by Quaker.

Sources of bias Firm incentives to engage in scientific research can create two sources of bias: 1) firms conduct research on only certain topics that are likely to showcase their

[3]For example, Kellogg's made the claim "Based upon independent clinical research, kids who ate Frosted Mini-Wheats cereal for breakfast had up to 18% better attentiveness three hours after breakfast than kids who ate no breakfast". The FTC investigated this claim finding no support for scientific evidence resulting in a complaint filed against the company (FTC File No. 0823145).

[4]https://www.quakeroats.com/about-quaker-oats/quaker-oats-center-of-excellence

products in a positive light (e.g. conduct research on heart-healthiness of a grain, but not on how high temperatures reduce beneficial properties of the grain) and 2) conditional on a given topic, industry-funded research finds/reports more favorable outcomes compared to non-industry funded research.

Definition: Unbiased Rating In what follows, the non-industry rating of an article will be considered to be the true unbiased rating. Note that non-industry articles might suffer from their own sources of bias (e.g. non-industry academics might conduct research on what are known to be controversial or negative topics). Bes-Rastrollo et al. (2013) referring to such biases by non-industry researchers state that "In an ideal world free from such biases, a perfect consistency between studies with different sources of funding would be expected" which will be the stance taken in this paper as well. Any differences between industry and non-industry funded research will be of interest.

Most sources of bias are subsumed in the above two main sources of bias. For example, publication bias – the suppression of null results either by authors or by the editorial process in either industry- or non-industry- funded articles – is captured in the broad category of the second kind of bias (conditional on a topic, one source reports more favorable outcomes than the other).

Definition: Topic I use keywords to define a topic. Web of Science records two types of keywords: author-provided keywords and Keywords Plus™. Keywords in Keywords Plus are generated by an algorithm unique to Clarivate Analytics databases and cover an article's cited works. According to Web of Science, Keywords Plus are index terms that frequently appear in the titles of an article's references (Garfield 1990).

One might worry that author-provided keywords are an endogenous construct, with authors choosing to define the 4-5 most related keywords. For example, articles that are industry-funded might choose keywords to appear more relevant or to appear in searches where most non-industry-funded articles showcase negative outcomes. Using keywords from Keywords Plus suffers less from this selection concern[5].

Furthermore, Keywords Plus terms have been shown to be more comprehensive and general than author keywords (Zhang et al. 2016; Garfield 1990). Finally, Keywords Plus has the added benefit of consistency when describing a given topic. Author-keywords are less consistent (e.g. dietary fibre and dietary fiber appear as two distinct keywords).

[5]One could argue that authors select who they cite. In such a case, defining a topic would become next to impossible because even if index terms were inferred from the text of the article, one could argue that the words in the article have been selectively chosen to showcase certain aspects.

In what follows, I therefore use Keywords Plus terms to define a topic. However, the results are robust to using author-provided keywords. One caveat is that a topic is defined by one keyword. As more keywords are included to define a topic the uniqueness of that topic will increase making it harder to find comparable industry- and non-industry- funded articles within a given topic.

In the next two subsections, I use the ratings data outlined in Section 2 to understand the source of bias that may exist in the food sciences. Specifically, using aggregate topic-level data, I determine whether the bias is driven by 1) selection on topics and/or 2) industry-funded research reporting more positive outcomes conditional on a topic.

3.1 Selection on topics

If firms are conducting research on selective topics, selecting on topics where outcomes are likely to be positive, we should see a positive correlation between industry-presence in a given topic and the true rating of that topic. On the other hand, if firms are trying to enhance their credibility via science, we should observe a fairly uniform distribution (similar to non-industry) in topics chosen.

The regression, specified in Equation 1, tests for any correlation between industry-presence and the true rating, $s \in \{+1, 0, -1, Unrelated\}$, of that topic. In this analysis a topic is defined by a keyword. An observation is at the keyword-abstract-rating level.

$$\mathbb{1}(Ind_k) = \alpha + \sum_s \beta_s \mathbb{1}(rating_{k,aw} = s) + \varepsilon_{k,aw} \tag{1}$$

Here, $\mathbb{1}(Ind_k)$ equals 1 if at least one firm has conducted research in topic k and $rating_{k,aw}$ is worker w's rating of abstract a in keyword k. Only non-industry abstracts and their corresponding ratings are included. Ratings can be positive (+1), negative (-1), neutral(0) or unrelated to health. A positive $\beta_{s=+1}$ would indicate that industry is more likely to be present in keywords that on average have positive true ratings. Because an article can have multiple keywords associated with it, the same article will be present multiple times. The regression, therefore, clusters at the keyword and article level. Articles that have no keywords associated with them and keywords with only one research article are excluded from this analysis.

Using a logit model to analyze this regression, Figure 2a reports the marginal effects. There are two noteworthy findings here. First, industry is no more likely than non-industry to work on positive vs. negative vs. neutral topics. Second, industry is 2% less likely to work on unrelated topics.

Taken together, this analysis shows that industry is not selecting on topics with likely positive outcomes, but is more likely to work on application-oriented (i.e., not unrelated to health) topics.

3.2 Bias toward reporting positive outcomes conditional on a topic

If industry-funded research is biased, conditional on a topic, we should see a skew towards more positive ratings for industry-funded articles. Figure 2b plots the density distribution of the ratings across topics by industry/non-industry funding suggesting industry-funded articles are positively skewed. The mean of the industry ratings across topics (dotted vertical line in Figure 2b) is greater than the mean of the non-industry ratings (solid vertical line) and this difference is statistically significant (t-statistic=6.32).

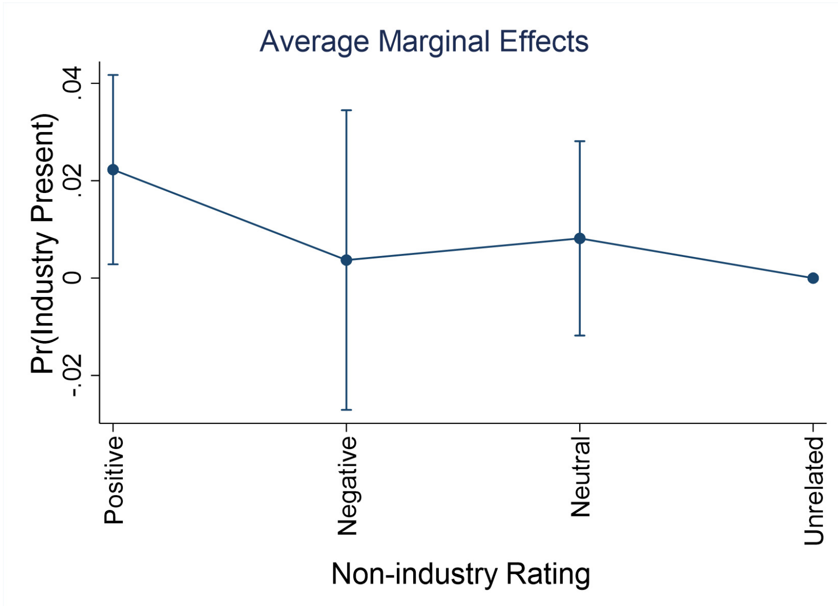
These two descriptive analyses suggest that industry is not selecting on topics with likely positive outcomes, but that conditional on a topic industry reports/finds more favorable outcomes than non-industry. This finding informs the analysis conducted in the next section, where I estimate regressions of the form:

$$y_a = \alpha + \beta Ind_a + \mathbf{A}_a + \varepsilon_a \tag{2}$$

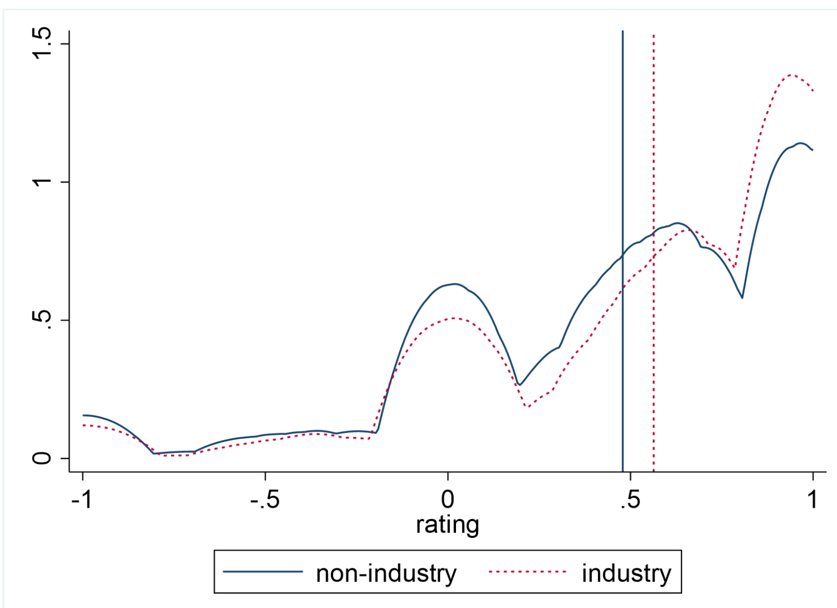
where y_a is the dependent variable pertaining to abstract a, $Ind_a = 1$ if the funding source for abstract a contains an industry participant, and $\mathbf{A}_a$ is a vector of controls. This linear regression model is appropriate in this setting because, although each article is rated as positive(+1), negative(-1) or neutral(0), the average rating per article is a continuous variable (e.g. 0.2, -0.8). A logit model with binary outcomes (not-positive versus positive) is presented in Appendix B as a robustness check.

Figure 2: Selection on topics vs. Bias conditional on a topic

(a) Industry-funded research presence across topics



(b) Density Plot of Average Ratings by funding source



Note: Figure 2a plots the results of regressing industry presence within a keyword on non-industry (unbiased) ratings within that topic N is 81,995 consisting of 2,525 titles across 2,106 keywords with 5 worker ratings per title. Keywords with only one article are omitted. An article can be present across multiple keywords. Standard errors are clustered at the keyword and article level. Appropriate sampling weights per food group are used. Figure 2b plots ratings across titles and workers aggregated to the keyword level by funding source. Articles with unrelated ratings are ignored for this analysis. There are 2,158 keywords in the industry group and 4,231 keywords in the non-industry group. Vertical lines indicate means of each distribution

4 Empirical Analysis

This section quantifies the extent of the bias by examining ratings at the more disaggregate article-level and with appropriate controls. To ensure comparison is within articles on the same research topic, I conduct analysis by grouping articles by their keyword. This within-keyword analysis enables comparison of industry-funded and non-industry funded articles within a research topic. An observation is, therefore, at the keyword-abstract-rating level. Because each article is associated with multiple keywords, the same article will be present multiple times. I therefore cluster standard errors at the keyword level and the article level (two-way clustering). Equation 3 specifies the regression:

$$y_{k,aw,gy} = \beta Ind_a + \gamma_1 Unrelated_{aw} + \gamma_2 Ind_a \times Unrelated_{aw} + \alpha_g + \alpha_y + \alpha_k + \varepsilon_{k,aw,gy} \quad (3)$$

where $y_{k,aw,gy}$ is worker w's rating of abstract a in the whole grain food group g published in year y and is associated with keyword k, and $Ind_a = 1$ if the funding source contains an industry participant. The coefficient β is the effect attributable to industry funding. Because "unrelated" ratings can be fundamentally different, equation 3 allows for such differences by absorbing such unrelated abstract ratings into a separate coefficient γ_1. The interaction term γ_2 further allows for "unrelated" ratings to differ by industry versus non-industry classification. The additional fixed effects α_g, α_y, and α_k correspond to the whole grain food group, year of publication, and keyword, respectively. Keyword fixed effects help allow for differences in topics chosen by industry versus non-industry. For example, if industry-funded articles focus on cardiovascular-related health outcomes but no such research exists for non-industry articles, such a selection into research categories can result in a spurious effect of industry on abstract rating. Similarly, inclusion of keywords with only non-industry participants might deflate the rating of non-industry, thus artificially inflating the effect of industry on rating. Controlling for keywords helps enable a comparison among similar research articles.

Table 5 shows the results of this regression with an increasing number of controls. Because disproportionate sampling was used, the regression accounts for the sampling probability by using the appropriate sampling weights per group. The results across various specifications are consistent and show industry-funded articles are on average 0.083 (β from column (4), in Table 5), more positive. To understand the magnitude of this effect, the range of the rating scale is 2 (-1 to +1), which implies that the degree of bias is fairly small at 4.1% (0.083/2) although significant.

Table 5: Average Rating of Industry-Funded Articles: All Whole Grains

Abstract Rating		(1)		(2)		(3)		(4)	
		coeff	t-stat	coeff	t-stat	coeff	t-stat	coeff	t-stat
Industry	β	**0.099**	**3.58**	**0.101**	**3.78**	**0.104**	**3.9**	**0.083**	**3.34**
Magnitude		5.0%		5.0%		5.2%		4.1%	
No. obs		137,365		137,365		135,045		135,045	
No. keywords		7,035		7,035		6,939		6,939	
No. abstracts		3,447		3,447		3,386		3,386	
Fixed Effects									
Food Group				Y		Y		Y	
Year of publication						Y		Y	
Keyword								Y	
Cluster		keyword, abstract							

Note: The table presents regressions of abstract ratings on funding source for all whole grains. Data post-2007 only used. Standard errors are clustered two-way over abstracts and keywords. The magnitude is computed by dividing β by 2 (the range of the rating scale).

4.1 Ratings by whole grain: Heterogeneous effects

To examine whether these effects differ by type of whole grain, I interact the coefficients β, γ_1, and γ_2 with each whole grain to obtain grain-specific coefficients. Specifically, I estimate the following regression equation:

$$\begin{aligned} y_{k,aw,gy} = \quad & \beta_g Ind_a \times \mathbb{1}\left(a_g\right) + \gamma_{1g} Unrelated_{aw} \times \mathbb{1}\left(a_g\right) + \gamma_{2g} Ind_a \times Unrelated_{aw} \times \mathbb{1}\left(a_g\right) \\ & + \alpha_g + \alpha_y + \alpha_k + \varepsilon_{k,aw,gy} \end{aligned} \tag{4}$$

where β_g, $\gamma_{1,g}$, and $\gamma_{2,g}$ are the grain-specific coefficients, and $\mathbb{1}\left(a_g\right)$ is an indicator that equals 1 if abstract a is in whole grain food group g. Table 6 reports β_g, the effect for industry-funded articles by each whole grain. The grains that have more industry-funded articles - corn, rice and wheat - see a higher impact of industry-funding on abstract positivity. In these whole grains, using estimates from Table 6, column (3), industry-funded articles are 0.16-0.19 (which on a scale with range 2 translates to 8%-9.5%) more positive.

Table 6: Industry-Funding impact: By Whole Grain Group

Abstract rating	(1)		(2)		(3)	
	coeff	t-stat	coeff	t-stat	coeff	t-stat
barley	-0.005	-0.07	-0.008	-0.10	-0.019	-0.23
buckwheat	0.012	0.05	0.014	0.06	0.083	0.32
corn	**0.187**	**2.46**	**0.192**	**2.51**	**0.161**	**2.37**
oatmeal	0.013	0.29	0.014	0.31	0.006	0.13
rice	**0.205**	**4.24**	**0.209**	**4.36**	**0.189**	**4.14**
rye	-0.132	-0.90	-0.125	-0.86	-0.133	-0.98
sorghum	0.081	0.80	0.086	0.84	0.077	0.67
triticale	0.928	2.80				
wheat	**0.183**	**2.97**	**0.190**	**3.15**	**0.162**	**2.94**
No. obs	132,245		130,025		130,025	
No. keywords	6,861		6,772		6,772	
No. abstracts	3,304		3,246		3,246	
Fixed effects						
Food Group	Y		Y		Y	
Year of publication			Y		Y	
Keyword					Y	
Cluster	keyword, abstract					

Note: The table reports β_g, the industry-funded coefficient for each whole grain. The whole grains bulgur, einkorn, farro, fonio, kaniwa, spelt and teff have no industry-funded observations. Grains with fewer than 10 industry-funded articles (amaranth, khorasan wheat, millet quinoa, wild rice) are excluded from this analysis. Highlighted rows indicate coefficients where at least two specifications have significant (at 95%) estimates. Data post-2007 only used. Standard errors are clustered at the abstract and keyword level.

4.2 Within author-topic comparison

To establish a more causal relationship, I examine outcomes when the same author conducts research on the same topic, with and without industry funding. Although assignment is not random, this analysis controls for author-specific correlations (e.g., industry may choose to partner with certain authors who work on topics with likely positive outcomes). In this analysis, an observation is at the author-topic-title-rating level. Standard errors are clustered multi-way across topics, authors and titles.

Restricting analysis to this subset, Table 7 Column 1 shows the effect continues to hold, although no longer statistically significant. I next further restrict attention to the grains with the heaviest funding: rice, wheat and corn which contribute most to the bias as documented in Section 4.1. Table 7 Column 2 shows not only a positive and significant effect for these grains, but also a larger magnitude of the effect at 0.240 which on a scale with range 2 (-1 to +1) translates to a bias of 12% (0.240/2).

Table 7: Industry-funding outcomes: Restricting attention to the same author-topic pair

Ratings	**(1)** All food groups		**(2)** Heaviest funding	
	coeff	t-stat	coeff	t-stat
Industry	**0.072**	**1.25**	**0.240**	**2.43**
Magnitude	3.6%		12.0%	
No. obs	9,170		4,355	
No. keywords	250		185	
No. abstracts	529		288	
No. authors	318		222	
Fixed Effects	food group year of publication keyword author			
Cluster	author, keyword, abstract			

Note: Column 1 reports results using data from all food groups. Column 2 restricts to grains with heaviest funding: rice, wheat and corn.

5 Determinants of the positivity bias

The results documented above suggest a significant, although small, bias of industry-funded research. Two main explanations for such a bias exist: 1) Industry-funded articles are of lower quality and/or report false or exaggerated claims and 2) Industry-funded articles are less scientifically written making interpretation easier for readers. I explore each possible explanation below finding little evidence for either explanation. Finally, I discuss one potential explanation that industry-funded articles are perhaps more application-oriented having sentences that describe how the findings can be applied to health outcomes. Because firms have incentives to disseminate their research to policy makers such additional words might be useful to aid in interpretation.

5.1 Are industry-funded articles of lower quality?

Metrics of an article's quality include the number of citations and the impact factor of the journal of publication. Moreover, abstracts containing the words *randomized* and *experiment* serve as indicators that the research was conducted using randomized designs, a gold-standard in the scientific literature.

Using these four metrics of quality, I run the following regression, specified in equation 5, at the research topic level. Within a research topic (keyword), I compare articles that are industry-funded vs. not, controlling for food group, year and topic fixed effects. Standard errors are clustered at the article and topic level.

$$y_{k,a,gy} = \beta Ind_a + \alpha_g + \alpha_y + \alpha_k + \varepsilon_{k,a,gy} \tag{5}$$

Here $y_{k,a,gy}$ is the dependent variable pertaining to abstract a in in the whole grain food group g published in year y and is associated with keyword k, and $Ind_a = 1$ if the funding source contains an industry participant. The coefficient β is the effect attributable to industry funding. The additional fixed effects α_g, α_y and α_k correspond to the whole grain food group, year of publication, and keyword, respectively. These fixed effects help control for factors such as the possibility that industry-funded articles might be more recent and hence have fewer citations.

Table 8, Panel a reports the results of the regressions. First, industry-funded articles receive 1.36 more citations than non-industry funded articles. Moreover, they publish in journals with slightly higher impact factors. The magnitude is an additional 0.25 points is significant at the 95% level. For reference, the average impact factor across all 144 journals is 2.44. Second, columns (3) and (4) of Table 8, Panel a show that industry-funded research is more likely to use randomized designs suggesting a higher quality of research.

Taken together, both the citation-level data and the abstract-text data suggest that industry-funded research is likely to be of higher quality. This could be driven by the access to funding resources industry has.

Table 8: Industry-funded Articles and Determinants of the Positivity Bias

Panel a. Metrics of Quality

	Publication quality				**Words in abstract**			
	(1)		(2)		(3)		(4)	
	citations		impact factor		randomized		randomized, experiment	
	coeff	t-stat	coeff	t-stat	coeff	t-stat	coeff	t-stat
Industry	**1.364**	**1.91**	**0.254**	**2.77**	**0.057**	**5.04**	**0.074**	**4.65**
Constant	13.28	46.06	2.90	11.55	0.021	8.63	0.110	20.66

Panel b. Scientific-ness and Abstract Length

	(1)		(2)		(3)		(4)	
	Scientific-ness		Total Words		Total Words w/o numbers		Target Words	
	coeff	t-stat	coeff	t-stat	coeff	t-stat	coeff	t-stat
Industry	**0.003**	**1.02**	**13.93**	**5.49**	**12.01**	**5.14**	**9.53**	**5.46**
Constant	0.167	135.46	197.74	207.57	188.01	208.89	140.73	207.06
No. obs	22,755							
No. keywords	2,681							
No. abstracts	3,366							
Fixed effects	food group year of publication keyword							

Note: Panel a presents regression results of various quality metrics of an article on whether that article was industry-funded. All standard errors are clustered at the keyword and article level. Not all journals have impact factors available resulting in a smaller subset, N=19,180, for the impact factor analysis. For the impact factor analysis, standard errors are clustered at an additional journal level because there is one impact factor per journal. Panel b presents regression results of various abstract-text metrics of an article on whether that article was industry-funded. The dependant variables in Column (1), Scientific-ness, is the ratio of scientific to common words in an abstract, Column (2), Total Words, is a count of all words in an abstract, Column (3), Total Words without Numbers, is a count of all words excluding numbers, and Column (4), Target Words, is a count of all words excluding articles, pronouns etc. All standard errors are clustered at the keyword and article level.

5.2 Are industry-funded articles less scientifically written?

If industry-funded articles are less scientifically written, a reader might be more easily able to interpret the results resulting in positive ratings. A scientifically-dense article, on the other hand, might be more likely to receive an unrelated or neutral rating. I also examine differences in the length of abstracts to understand if there are any underlying differences.

Using a list consisting of the top 60,000 most frequently used words in written English articles (Word Frequency Data), I count the number of words in an abstract that are present in this list. The remaining words are likely to be of a more scientific nature (e.g. Nonylphenol). Using this procedure I create, for each abstract, a ratio of scientific to total words (net of articles, conjunctions, determiners, prepositions, pronouns and interjections). I also count the total number of words in an abstract.

Using these metrics I run the same regression as specified in equation 5. Table 8, Panel b reports the results of this regression. Column (1) shows that industry-funded articles are no different than non-industry-funded articles in the scientific-ness of their abstracts. However, columns (2), (3) and (4) – all of which use variants of the total word count of an abstract – show that industry-funded abstracts have on average 10-14 more words.

One potential explanation for this finding is that all else equal, industry-funded abstracts explain their findings and suggest avenues for application. Taking as an example, two articles by the same author-topic pair that differ only in the funding source, the abstract from the industry-funded article contains additional text towards the end of the abstract stating "Therefore reduction in the harvest/drying interval would be essential to assure product quality and safety and minimize potential health hazards." suggesting that industry-funded research abstracts provide better links between the findings and the application of those findings to health outcomes.

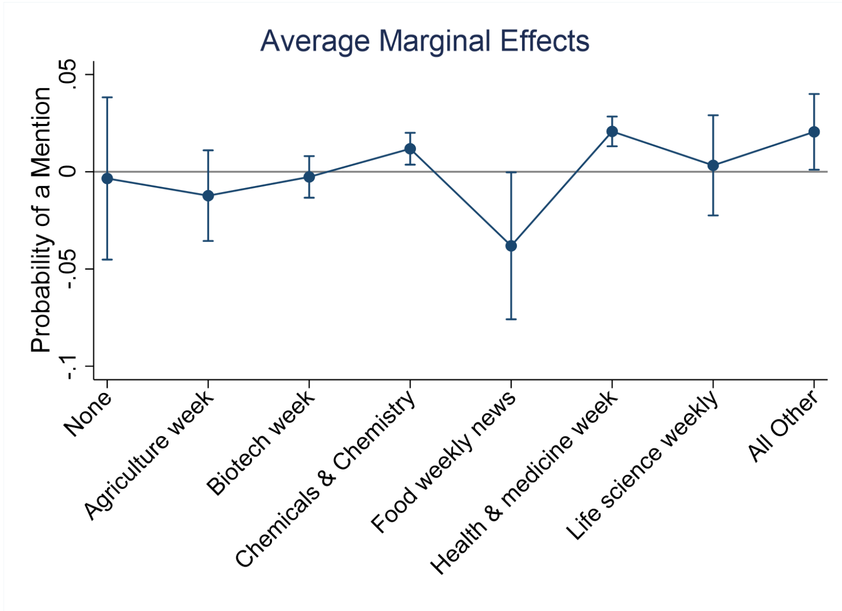
6 News mentions

Analyzing mentions in various trade press and industry newsletters, I find evidence suggesting that industry-funded articles are more likely to receive more trade press coverage. To understand if certain newsletters are more or less likely to feature industry articles, I analyze the various news sources using a multinomial logit model specified in Equation 6:

$$p\left(y_{k,a,gy} \text{ in } n\right) = \frac{exp\left(\alpha_n + \beta_n Ind_a\right)}{\sum_{c=1}^{C} exp\left(\alpha_c + B_c Ind_a\right)} \tag{6}$$

where $p\left(y_{k,a,gy} \text{ in } n\right)$ is the probability that research article a is mentioned in news outlet n, β_n measures the additional propensity of an industry-funded article being mentioned in news-outlet n. Here C is the set of possible news outlets and includes the top six news outlets with the rest grouped into All Other.

Figure 3 plots the marginal effect of industry-funding on the probability that the article was mentioned in that news outlet. The analysis reveals that certain newsletters (*Chemicals & Chemistry*, *Health & Medicine Week* and *All Other*) are more likely to feature industry-funded articles. To the extent that industry-funded articles are being featured in certain trade magazines/newsletters, it is suggestive of a channel to reach influencers (e.g. policy makers, nutrition professionals) who in turn have the ear of the consumer.



Note: Figure plots the marginal effect of industry-funding on probability that the article was mentioned in a newsletter. Specification uses multinomial logit regression. Standard errors are clustered at the keyword and article level. N is 16,329 consisting of 7,737 keywords and 3,188 titles.

Figure 3: Probability of industry-funded research receiving a mention in a newsletter

7 Conclusion

This paper finds evidence that industry-funded articles are more likely to document positive findings and garner media mentions in various research-centric outlets. Although research has documented the role of industry in emphasizing the less harmful effects of its products, this paper aims to document existence of bias in healthy food groups finding evidence for a small degree of bias.

This paper finds little evidence for strategic selection of research topics, i.e. industry is no more likely than non-industry to work on topics with likely positive outcomes. However, industry-funded research is likely to be more application-oriented. Within a research topic, industry-funded research reports more positive outcomes but these are not likely to be false or unsubstantiated claims as evidenced by the higher quality of research conducted by industry. One potential explanation is that industry explains their findings aiding in interpretation and application, perhaps a more relevant goal for industry relative to academics.

The data used in this paper was from 2007 to 2020. With more years' worth of data and as more industry-funded articles become available, exploring time trends in the data will help us understand whether such research shapes consumer preferences. For example, lately amaranth and quinoa are receiving more attention from consumers and firms alike.

Understanding whether industry- or independent- research shapes such trends would be an avenue for future research.

Finally, using the text of research articles to determine study design parameters such as sample size can shed more light on the reasons for the existence of the bias documented in this paper.

References

[1] Bekelman JE, Li Y, Gross CP. (2003), "Scope and impact of financial conflicts of interest in biomedical research: a systematic review", JAMA, 289 (4):454–69.

[2] Belluz, Julia (2018), "Nutrition research is deeply biased by food companies. A new book explains why.", Vox, Nov 11 2018, https://www.vox.com/2018/10/31/18037756/superfoods-food-science-marion-nestle-book, Accessed February 2021.

[3] Bes-Rastrollo M, Schulze MB, Ruiz-Canela M, Martinez-Gonzalez MA. (2013), "Financial Conflicts of Interest and Reporting Bias Regarding the Association between Sugar-Sweetened Beverages and Weight Gain: A Systematic Review of Systematic Reviews", PLoS Medicine, 10(12).

[4] Burke, S. J., Milberg, S. J., & Moe, W. W. (1997), "Displaying Common but Previously Neglected Health Claims on Product Labels: Understanding Competitive Advantages, Deception, and Education", Journal of Public Policy & Marketing, 16(2), 242–255.

[5] Cooper RJ, Gupta M, Wilkes MS, Hoffman JR. (2006), "Conflict of interest disclosure policies and practices in peer-reviewed biomedical journals", Journal of General Internal Medicine,21(12), 1248–1252.

[6] Dellavigna S, Hermle J. (2017), "Does Conflict of Interest Lead to Biased Coverage? Evidence from Movie Reviews", The Review of Economic Studies, 84(4), 1510–1550.

[7] Diels J, Cunha M, Manaia C, Sabugosa-Madeira B, Silva M (2011), "Association of financial or professional conflict of interest to research outcomes on health risks or nutritional assessment studies of genetically modified products", Food Policy,36(2), 197-203.

[8] Dow Jones & Co, Dow Jones Reuters Business Interactive LLC, & Reuters ltd. (2001). Factiva. [New York]: Dow Jones Reuters Buisness Interactive LLC.

[9] Dranove D, Jin GZ (2010), "Quality Disclosure and Certification: Theory and Practice", Journal of Economic Literature, 48, 935–63.

[10] Food Engineering (2019), "2019 Top 100 Food & Beverage Companies", https://www.foodengineeringmag.com/2019-top-100-food-beverage-companies, Accessed June 2020.

[11] Food Manufacture Directory, https://www.foodmanufacturedirectory.co.uk/, Accessed June 2020.

[12] Garfield, E (1990), "Keywords Plus®: ISI's breakthrough retrieval method. Part 1. Expanding your searching power on Current Contents on Diskette", Current Contents, 1(32), 5–9.

[13] Hastak, M., & Mazis, M. B. (2011), "Deception by Implication: A Typology of Truthful but Misleading Advertising and Labeling Claims", Journal of Public Policy & Marketing, 30(2), 157–167.

[14] Institute for Scientific Information. (1997), "Web of science", Philadelphia, PA: Institute for Scientific Information.

[15] Kearns CE, Schmidt LA, Glantz SA (2016), "Sugar Industry and Coronary Heart Disease Research: A Historical Analysis of Internal Industry Documents", JAMA Internal Medicine, 176(11), 1680–1685.

[16] Krimsky S, Rothenberg LS (2001), "Conflict of interest policies in science and medical journals: editorial practices and author disclosures", Science and Engineering Ethics, 7, 205–18.

[17] Lesser LI, Ebbeling CB, Goozner M, Wypij D, Ludwig DS (2007), "Relationship between funding source and conclusion among nutrition-related scientific articles", PLoS Medicine, 4(1), 41-46.

[18] Levine J, Gussow JD, Hastings D, Eccher A. (2003), "Authors' financial relationships with the food and beverage industry and their published positions on the fat substitute olestra", American Journal of Public Health, 93(4), 664-669.

[19] Lexchin J, Bero LA, Djulbegovic B, Clark O. (2003), "Pharmaceutical industry sponsorship and research outcome and quality: systematic review", BMJ, 326, 1167–70.

[20] Massougbodji J, Le Bodo Y, Fratu R, DeWals P. (2014), "Reviews examining sugar-sweetened beverages and body weight: correlates of their quality and conclusions", The American Journal of Clinical Nutrition ,99(5), 1096-1104.

[21] Moorman, C. (1998), "Market-Level Effects of Information: Competitive Responses and Consumer Dynamics", Journal of Marketing Research, 35(1), 82-98.

[22] Myers EF, Parrott JS, Cummins DS, Splett P. (2011), "Funding source and research report quality in nutrition practice-related research", PLoS One. 6(12), e28437.

[23] Nestle, Marion (2018), "Unsavory Truth: How Food Companies Skew the Science of What We Eat", Basic Books

[24] Nestle, Marion (2020), "Annals of food marketing: pistachios have amino acids (duh)!", Food Politics, July 6 2020, https://www.foodpolitics.com/2020/07/23390/ Accessed September 2020.

[25] Nkansah N, Nguyen T, Iraninezhad H, Bero L. (2009), "Randomized trials assessing calcium supplementation in healthy children: relationship between industry sponsorship and study outcomes", Public Health Nutrition, 12(10), 1931-1937.

[26] Oostrom, T. (2021), "Funding of Clinical Trials and Reported Drug Efficacy", working paper.

[27] Piazza T. (2010), "Fundamentals of Applied sampling", in Marsen PV, Wright JD (eds). Handbook of Survey Research (2nd Edition), Emerald Group Publishing, Bingley, UK.

[28] Reuter, J. and Zitzewitz, E. (2006), "Do Ads Influence Editors? Advertising and Bias in the Financial Media", The Quarterly Journal of Economics, 121, 197–227.

[29] ReferenceUSA (2020), Companies under SIC Code 20, Retrieved from ReferenceUSA database.

[30] Solin, Paul (2015), "Company Name Processor written in Python", https://github.com/psolin/cleanco, Accessed June 2020.

[31] Sismondo S. (2008a), "Pharmaceutical company funding and its consequences: A qualitative systematic review", Contemporary Clinical Trials, 29, 109–113.

[32] Sismondo S. (2008b), "How pharmaceutical company funding affects clinical trial results: causal structures and responses", Social Science & Medicine, 66, 1909-1914.

[33] Vartanian, L.R., Schwartz MB, and Brownell KD. (2007), "Effects of Soft DrinkConsumption on Nutrition and Health: A Systematic Review and Meta-Analysis", American Journal of Public Health 97, 667-675.

[34] Wilde P, Morgan E, Roberts J, Schpok A, Wilson T. (2012), "Relationship between funding sources and outcomes of obesity-related research", Physiology & Behavior, 107(1), 172-175.

[35] Word Frequencey Data, https://www.wordfrequency.info/intro.asp, Accessed February 2021.

[36] Zhang, J., Q. Yu, F. Zheng, C. Long, Z. Lu and Z. Duan (2016), “Comparing Keywords Plus of WOS and Author Keywords: A Case Study of Patient Adherence Research”, Journal of the Association for Information Science and Technology, 67(4): 967-972.

A Case study: Oats

Of all oatmeal abstracts, 13% are industry-funded. Moreover, oatmeal is monopolized by Quaker Oats (Pepsi) not only in market share, but also in industry funding. Among all industry-funded abstracts, Quaker Oats contributes a disproportionate 32%, with the next firm (General Mills) contributing only 7%. This pattern suggests Quaker might have incentives to convince consumers its product is healthy.

To examine if industry-funded articles are more positive, I run the following descriptive regression for the dependent variable *ratings*:

$$y_{aw} = \alpha + \beta Ind_a + \varepsilon_{aw} \tag{7}$$

where y_{aw} is worker w's rating for abstract a in oatmeal numerically coded as 1 for positive ratings, 0 for neutral or unrelated ratings, and -1 for negative ratings. $Ind_a = 1$ if the funding source for abstract a contains an industry participant such as Quaker. The coefficient α is a constant that represents the non-industry average of the dependent variable and β is the effect attributable to industry funding. All standard errors are clustered at the abstract level.

Table 9 presents the results of this regression showing that industry-funded articles are more positive. Industry-funded articles are 0.088 (β from Table 9) more positive. Because the rating scale ranges from -1 to +1, this implies a bias of 4.4% (0.088/2). However, this finding could also occur if industry-funded articles focus more on certain types of research that non-industry might choose not to focus on. In the empirical analysis section of the paper, where I use all whole grains, I therefore conduct analysis at the keyword level, comparing all abstracts associated with the same keyword along with other controls.

In the above regression, "unrelated" ratings are assumed to be the same as "neutral" ratings (i.e., are assigned a value of 0). Because "unrelated" ratings can be fundamentally different, I allow for such differences in the main empirical analysis where I absorb such unrelated ratings into a separate coefficient.

Table 9: Industry Funding and Abstract Ratings: Oats

Average Rating		coeff	t-stat
Industry	$\boldsymbol{\beta}$	**0.088**	**2.34**
Constant	α	0.379	29.74
No. of obs		4,250	
No. of abstracts		850	

Note: The table presents the regression of abstract ratings on funding source for the food group oatmeal. Unrelated ratings are coded as neutral (numerical rating of 0). Data post-2007 only used. Standard errors are clustered at the abstract level.

B Logit model of abstract ratings

I estimate a logit model where ratings are converted to a binary outcome of positive versus not-positive. The logit model is specified in equation 8:

$$p\left(y_{k,aw,gy}=1\right)=\frac{exp\left(\alpha+\beta Ind_a+\alpha_g+\alpha_y\right)}{1+exp\left(\alpha+\beta Ind_a+\alpha_g+\alpha_y\right)} \tag{8}$$

where $p\left(y_{k,aw,gy}=1\right)$ if abstract a is rated as positive and $p\left(y_{k,aw,gy}=0\right)$ if the abstract is rated as negative or neutral (i.e., not positive) by worker w. $Ind_a=1$ if the funding source contains an industry participant. The coefficient α is the effect attributable to non-industry articles, and β is the effect attributable to industry funding. The additional fixed effects α_g and α_y correspond to the whole grain food group and year of publication respectively. Due to the large number of keywords, keyword fixed effects are not estimated in this logit specification. In this estimation, I exclude those abstract-ratings that are rated as unrelated to health outcomes.

Table 10 presents the marginal effect of industry-presence from the logit models with an increasing set of controls. Using the estimates from Table 10, column (3), the results indicate that industry-funded articles are 7.3% more likely to be positive (as opposed to negative or neutral).

Table 10: Rating of Industry-Funded Articles: Marginal Effects

Abstract Rating	(1)		(2)		(3)	
Positive vs. Not Positive	coeff	t-stat	coeff	t-stat	coeff	t-stat
Industry	**0.062**	**2.94**	**0.074**	**3.63**	**0.077**	**3.75**
No. obs	52,414		52,299		51,740	
No. keywords	5,258		5,253		5,196	
No. abstracts	2,289		2,286		2,253	
Fixed Effects						
Food Group			Y		Y	
Year of publication					Y	
Cluster	keyword, abstract					

Note: The table presents marginal effects estimated from a logit model of ratings (coded as 1: positive, 0: negative/neutral) on funding source. Abstracts rated as unrelated by workers are dropped. Data from all whole grains used. Data post-2007 only used. Standard errors are clustered at the abstract and keyword level.