Abstract

We propose a theoretical model of advertising content, based on an information-persuasion trade-off. In an advertisement, brands must allocate available time between imparting objective information about the product’s attributes and utilizing other persuasive methods. The theoretical model gives rise to a structural ordered probit model, which we estimate with data from the OTC analgesics industry. We test four predictions of the model. Stronger vertical differentiation is positively associated with the delivery of more product information in a brand’s advertisements: brands with higher levels of quality (along each of the quality dimensions) include more information cues. Comparative advertisements contain significantly more product information than non-comparative advertisements. Brands with higher market shares and brands competing against generic substitutes with higher market shares have less information content. The method of measuring and analyzing information content of advertising improves upon existing techniques.

Keywords: information content, advertising, information-persuasion trade-off, content analysis
Information Content of Advertising: Theory and Empirical Evidence

How much information brands choose to disclose in advertisements is a question of considerable theoretical and empirical debate. Recent research in marketing and economics provides some theoretical predictions on the relationship between market structure or firm size and the amount of information transmitted (Anderson and Renault 2009; Guo and Zhao 2009; Sun 2010). The empirical work on advertising content is split into two camps, with essentially no overlap. The first camp uses Resnik and Stern’s (1977; henceforth RS) methodology to analyze content (for a summary, see Abernethy and Franke 1996). The second camp treats advertising content as a choice variable (Anderson et al. 2011; Bertrand et al. 2010; Liaukonyte 2011). This paper develops an information-persuasion trade-off theory model, where the “persuasive” content of an ad is interpreted here as the content that is not objective information, and also enables us to bridge the gap between the two camps by applying the theory model to the novel data.

The key idea of this paper is that there exists an optimal amount of information to be included in an advertisement. Advertisements that provide too little objective information about the brand arguably waste the opportunity to sufficiently convince prospective consumers to buy it (Jacoby 1977). Conversely, those that provide too much information may crowd the ad message and lead to information overload for the consumer (Chervany and Dickson 1974; Peters, Wedel and Zhang 2007). Factors such as motivation and the ability to process information (Cacioppo and Petty 1985; MacInnis and Jaworski 1989) mediate individual responses to advertising. Therefore, the complexity of advertising, including too many information cues, can create attention wear-out (Pieters, Warlop and Wedel 2002), which suggests that an optimal degree of information content exists. The optimal amount of information content may vary systematically across brands and may be partially explained by observable factors such as brand type, brand size, suitability of various combinations of information, and recent news about the product.

We formulate a theoretical model in which firms decide how much objective information to include in an advertisement. The model allows for a trade-off between objective information and other persuasion, subject to random factors intrinsic to specific ads. The amount of time spent on information—as approximated by the number of information cues—has systematic and random components. This approach allows us to provide a theoretical underpinning for an ordered probit model of the number of information cues in advertisements. Furthermore, the theoretical model also yields predictions on the relationship between information content and various observable performance measures (such as market share). The strength of the theoretical model is that it is intuitive, stylized, and simple, and provides micro-foundations for the application of the ordered probit model to content analysis, but it is limited to a monopoly context. The extension to oligopoly is left for future enrichment.
The empirical component of this paper examines the relationship between market variables and the information content of advertising. First, we classify and fully measure different types of advertising content, as well as the distribution of information cues, within an entire industry. Second, we relate the extent of information disclosure to the market share of a brand, its core vertical characteristics, and the share of the generic substitute. To accomplish this, we formulate four hypotheses about the influence of the type of advertising, brand vertical characteristics, market share, and generic substitute market size on the information content of advertising. The empirical results substantiate intuitive patterns of advertising content and quantify statistical regularities.

In line with our theoretical results of a trade-off between the informative and persuasive components of advertising content, we find empirical support for the hypothesis that stronger vertical differentiation is positively associated with more information. We also find that comparative advertisements contain significantly more information than non-comparative advertisements, and that larger brands and a higher market share of the generic version of a brand are both associated with less informative ads. Significant estimation bias results from not controlling for the endogeneity of the decision to use comparative advertising and from the endogeneity of market share. Finally, we show that the results are largely robust to using the RS information content definition; however, the RS methodology does not fully account for all of the aspects of information that are important to consumers.

**Advertising Content and the Information-Persuasion Tradeoff**

**Background and Definitions**

In general, information in advertisements can be (1) about the brand, brand attributes, benefits, users, or usage situation; (2) cognitive, emotional or subconscious; or (3) contextual information, including consumers’ past experiences (MacInnis, Moorman, and Jaworski 2001; Vakratsas and Ambler 1999). In our theoretical model and then later in the empirical application, we focus on the first aspect of brand information (tractable, objective information cues) and treat cognitive, emotional, and contextual information as the intractable, subjective element of an ad. We define this subjective element as the *persuasive component*. Figure 1 summarizes our approach. The shaded area represents the tractable informative component of an ad, whereas the non-shaded area represents the intractable (to us) persuasive component.

Contrary to the classic content analysis, which determines the fraction of advertisements that fall into each category given a number of cues and uses univariate analysis to compare scenarios, we use an

1 The persuasive component can also be tractable in well-defined studies (e.g., Maheswaran and Joan Meyers-Levy 1990), but it is outside of the scope of this paper.
ordered probit model to study the *determinants* of the distribution of cues. The ordered probit is directly derived from a theoretical model that describes the equilibrium choice of information content as a function of the exogenous variables. Our model is based on an intuitive relationship: a greater amount of information content is associated with a larger number of cues.

It is difficult to derive clear-cut testable hypotheses based on existing theory. First, most economic and marketing theories of advertising do not address the content of advertising (Anderson and Renault 2006, 2009) and most models deal with monopolies: even oligopoly models usually assume that firms are symmetric. Second, standard advertising models typically address only one type of advertising (persuasive or informative), while any given advertisement likely incorporates several components simultaneously.

**Figure 1. Determinants of Information Disclosure**

---

**The Model**

Our theoretical model distinguishes between two types of advertising content: information and persuasion. The firm faces a trade-off: given the limited amount of time available in an ad, the firm must decide how much of that time to allocate to providing information and how much of it to allocate to persuading consumers to buy the brand through the use of channels other than objective information. We assume that the persuasive power of an advertisement depends on the number of seconds devoted to the persuasive component, and a random term which reflects the idiosyncratic features of a particular advertisement. Let the persuasive power of the ad be

\[ P(s, \varepsilon) = (\bar{p} - \varepsilon)s \]
Figure 2. Tradeoff Between Informative and Persuasive Content

Panel A

- Information Benefit
- Persuasion Benefit
- Incremental Benefit of 4th Information Cue
- Incremental Benefit of 5th Information Cue
- Opportunity Cost of 5th Information Cue
- Time Spent on Informative Cues
- Optimal Number of Cues
- Length of Time for 1 Cue
- Length of an Ad
- Persuasive Content

Panel B

- Incremental Information Benefit
- Incremental Persuasion Benefit
- $\overline{p} - \epsilon_1$
- $\overline{p} - \epsilon_2$
- $\overline{p} - \epsilon_3$
- $\frac{\partial P}{\partial s}(s, \epsilon) = \overline{p} - \epsilon$
- Time Spent on Informative Cues
- Optimal Number of Cues
- Persuasive Content

$P(s, \epsilon) = s(\overline{p} - \epsilon)$
where \( s \) is the number of seconds spent on persuasion, \( \bar{p} \) is a constant and \( \epsilon \) is a random term.\(^1\) Panel A of Figure 2 illustrates the function \( P(s,\epsilon) \). The x-axis indicates the number of seconds used for persuasion. As more time is allocated to persuasion, the function \( P(s,\epsilon) \) increases from right to left. The y-axis shows the total benefit from persuasion. The more time spent on persuasion, the larger the total benefit.

The marginal persuasion for an advertisement with \( s \) seconds of persuasion and a given draw of the random term \( \epsilon \) is therefore \( \bar{p} - \epsilon \). The persuasion function is linear, and therefore the marginal persuasion does not change with the share of the advertisement that is allocated to it. Panel B of Figure 2 illustrates the function \( \bar{p} - \epsilon \). The x-axis again indicates the number of seconds used for persuasion. Because \( \bar{p} - \epsilon \) does not change with \( s \), the function \( \bar{p} - \epsilon \) is parallel to the x-axis. The y-axis depicts the marginal benefit of persuasion. In the econometric model presented later, \( \bar{p} \) varies according to the observable features of the advertised brand (e.g., market share, observable quality, generic competition, etc.) and of the advertisement itself, such as whether it is a comparative or non-comparative advertisement.

Let each information cue take \( \bar{s} \) seconds to convey, so that if there are \( S \) seconds in the ad (i.e. \( n \) information cues are conveyed), there are \( s=S-n\bar{s} \) seconds of persuasion. Let \( I_i \) be the benefit of the \( i^{th} \) information cue \( i \), with \( i=1,...,n \). We rank the cues from the highest to the lowest information benefit for each given advertisement (and the ranking may differ across advertisements according to the particular theme of the ad). Clearly, the brand will choose to include the cues delivering the highest information benefit, i.e., those cues for which the values of \( I_i \) are the highest.

Because a given advertisement only has a limited amount of available time, \( S \), the brand must decide how much of that time to allocate to providing information vs. for persuasion. This trade-off and the total benefit of information are depicted in Panel A of Figure 2. Panel B of Figure 2 shows the marginal benefit of information, which is decreasing in the amount of information already provided. The marginal benefit of information is also a step function. The firm chooses \( s \) (or, alternatively, \( n \)) to maximize the sum of the total benefit of persuasion and the total benefit of information. Formally, the firm solves:

\[
\begin{align*}
\max_s \quad & \sum_{i=1}^{n} I_i + P(s,\epsilon) \\
\text{s.t.} \quad & n = \frac{S - s}{\bar{s}}.
\end{align*}
\]

Here \( D(\cdot) \) is an increasing function representing the firm’s demand as a function of an advertisement’s information and persuasion content. The solution to this optimization problem can be described by comparing the incremental benefit from adding a cue to the advertisement to the opportunity cost of reducing the time spent on persuasion. If the advertisement contains \( n-1 \) cues, then the extra benefit from

\(^1\) \( P(s,\epsilon) \) does not have to be linear in \( s \). It should be an increasing and concave function of \( s \). The linear specification simplifies the exposition significantly.
the $n^{th}$ cue is $I_n$. We can see this graphically in Panel A of Figure 2. There we see that $I_4$, the marginal benefit of the fourth information cue, is larger than $I_5$, the marginal benefit of the fifth information cue. The slope of the persuasion function, $\bar{p} - \varepsilon$, is such that:

$$I_5 < \bar{s}(\bar{p} - \varepsilon) < I_4.$$ 

This implies that the optimal number of cues depicted in Panel A of Figure 2 is four.

Panel B of Figure 2 shows this solution in a way that allows us to introduce our statistical model in a straightforward way. Define for all $I_i$ the value $\varepsilon_i$ such that

$$I_i = (\bar{p} - \varepsilon_i) \bar{s},$$

so that $\varepsilon_i$ is the threshold value of the random error such that the firm chooses to include at least $i$ information cues in the advertisement if $\varepsilon > \varepsilon_i$. In Figure 2, Panel A depicts $\varepsilon_4 < \varepsilon < \varepsilon_5$. Hence, in this particular illustration, the brand chooses to include four information cues in the advertisement.

The formulation above underpins our statistical model. Accounting for the fact that there can be no fewer than zero cues and no more than $C = S/\bar{s}$, we have the following mapping from the value of $\varepsilon$ to the number of cues, here denoted by $y$:

$$\begin{align*}
  y &= 0, & \text{for } \varepsilon \leq \varepsilon_1 \\
  & \vdots \\
  y &= i, & \text{for } \varepsilon_i < \varepsilon \leq \varepsilon_{i+1} \\
  & \vdots \\
  y &= n, & \text{for } \varepsilon > \varepsilon_n.
\end{align*}$$

The basic intuition of this statistical model is the following: when the (negative) random shock is very small ($\varepsilon < \varepsilon_1$)—which implies that an advertisement has ability to have very strong persuasive power (relative to the benefit of information in that particular advertisement)—then the firm has no incentive to include information cues. As the persuasive power of an advertisement decreases, the firm chooses to include more information cues.

From the specification above we can construct a probability distribution of observing the corresponding number of cues, where $F(\cdot)$ denotes the cumulative distribution of $\varepsilon$:

$$\begin{align*}
  y &= 0, \text{ with probability } F(\varepsilon_1) \\
  & \vdots \\
  y &= i, \text{ with probability } F(\varepsilon_{i+1}) - F(\varepsilon_i) \\
  & \vdots \\
  y &= n, \text{ with probability } 1 - F(\varepsilon_n).
\end{align*}$$
If $F(.)$ is normally distributed, the formulation corresponds to an ordered probit model. In particular, we show how the probability function in the model above can be written as the textbook version of the ordered probit, where $\varepsilon_i = \alpha_i - X\beta$ and therefore $F(\varepsilon_i) = \Phi(\alpha_i - X\beta)$. In this ordered probit specification the unobserved components, i.e. $\varepsilon$, are drawn from a normal distribution, and the cutoff values ($\alpha_i$) are such that the realization of a latent variable (explained component plus noise) lies within a range that corresponds to each specific number of cues. $\beta$ is a $K \times 1$ vector of parameters, and $X$ is a $K \times 1$ vector of observable features of the brand, which does not include a constant. For $\beta > 0$, an increase in $X$ will lower the threshold $\varepsilon_i$ which will in turn make it more desirable to add additional information cues. Therefore, for $\beta > 0$, larger $X$s are associated with more information content in an advertisement.

We can rewrite $F(\varepsilon_i) = \Phi(\alpha_i - X\beta)$. First, recall that $\varepsilon_i = \bar{p} - \frac{l_i}{\bar{s}}$. Here $\bar{p}$ is a variable, the value of which determines the benefit of persuasion content. Hence, we set $\alpha_i - X\beta = \bar{p} - \frac{l_i}{\bar{s}}$. It is useful to extract a constant from $-\frac{l_i}{\bar{s}}$ and to rewrite the equality as $\alpha_i - X\beta = \bar{p} - \bar{I} - \frac{l_i}{\bar{s}}$. Then $X\beta = \bar{I} - \bar{p}$ so that, consistent with the discussion in the preceding paragraph, the $X$ variables increase the benefit of information or, equivalently, decrease the benefit of persuasion. Second, we can define the cutoff value $\alpha_i = -\frac{l_i}{\bar{s}}$, which determines whether the firm is including $i$ or $i+1$ information cues. The econometric model does not identify the constant term $\bar{I}$ separately from the cutoffs $\alpha_i$, which we need to keep in mind when we interpret the cutoffs.

In our framework, the cutoffs have a clear structural interpretation. In particular:

$$\alpha_{i+1} - \alpha_i = \frac{l_i - l_{i+1}}{\bar{s}}.$$ 

Thus, conditional on $\bar{s}$ (which is unobserved), differences in the cutoffs provide information on differences in the information benefits of an additional cue. This observation enables us to take results from our econometric model and use them to draw a graph of the estimated cutoffs, and compare it directly with theoretical relationship depicted in Figure 2. We do this in the empirical section below.

We estimate this structural model and use the results to examine the relationships between the fundamental variables (e.g., a market share of a firm) and the number of information cues that a firm includes in an advertisement. Before discussing the relationships of interest, we state and prove the main theoretical result which drives the rest of our analysis and helps us formulate key hypotheses.

---

2 For a comprehensive review of ordered models, see Greene (1997) and Woolridge (2001).

3 For concreteness, we describe the random term as entering the persuasion power, but it could just as well enter the marginal information benefit described below. Or, indeed, the random term could enter on both sides and the analysis would then capture the net effect.
Lemma 1. An increase in $I_i$ stochastically increases the number of information cues.

Proof. From the analysis above, $i$ cues will be advertised if $\varepsilon \in (\varepsilon_i, \varepsilon_{i+1})$, where, as above, $\varepsilon_i = \bar{p} - \frac{I_i}{S}$. The corresponding probability of observing $i$ cues is $P_i = F(\varepsilon_{i+1}) - F(\varepsilon_{i})$, or

$$P_i = F(p - \frac{I_{i+1}}{S}) - F(p - \frac{I_i}{S}).$$

Suppose now that $I_i$ increases while retaining its position as the $i$th largest information benefit. Then $P_i$ increases at the expense of $P_{i-1}$, while all other probabilities remain unchanged. Hence, the number of cues increases stochastically. Now suppose that the increase in $I_i$ raises it to the $j^{th}$ highest cue, with $j < i$. Then each intervening cue is promoted so that the probability of observing at least that number of cues rises. The probability of observing $i+1$ cues or more stays the same, as does the probability of observing each number less than $j$. Again, the number of cues stochastically increases. QED

Vertical Characteristics

First, we consider intrinsic characteristics of a product. In our empirical application of the theoretical model, the examples of intrinsic attributes include but are not limited to: strength of pain relief, relative efficiency, safety, etc. In the industry analyzed in this paper, these variables are naturally exogenous to the information decision. This is because they depend on the medical properties of the active ingredients in analgesic pain relievers, which in turn are inflexibly regulated by the Federal Drug Administration (FDA). Therefore, the direction of causality can be clearly identified, and we can investigate how different locations in the product characteristics space are associated with information disclosure.

The driving idea here is that the information benefit to the brand from communicating a characteristic is increasing in the strength of the performance of the brand in that characteristic. Communicating a weak characteristic does not give as much incremental benefit as communicating a strong one. Characteristics which are important to consumers are more likely to be communicated, and additional strength in any particular cue will raise its relative information benefit and also make it more likely to be included in the advertisement.

Thus, using the result from Lemma 1 we formulate:

**H1**: Brands with higher levels of quality (on each of the quality dimensions for which we have data) will include more information cues in their advertisements.
**Comparative Advertising**

The second relationship that we study is the one between the decision to make a comparative claim and amount of information provided. There is no existing theory that tells us whether comparative advertisements should include more or less information than non-comparative advertisements.\(^4\) More information is conveyed by comparing two brands than just promoting a brand, and comparing relative performance provides a more precise and concrete reference point. On the other hand, an advertisement with comparative content is likely to have a weaker persuasion effect, even if the amount of time devoted to persuasion is the same. This might happen because mentioning the other brand dilutes the persuasion because it reiterates the existence of the rival brand. Previous research (Chou, Franke, and Wilcox 1987; Harmon, Razzouk, and Stern 1983) has found that comparative advertisements have more information, and we expect to find similar patterns.

We might expect a comparative advertisement to both increase the marginal information and decrease the marginal persuasion benefit of an advertisement. Both effects cause the number of cues to rise stochastically (see Figure 2). This leads us to formulate our second hypothesis:

\[ \textbf{H2: Comparative advertisements contain more information than non-comparative advertisements.} \]

**Market Share**

The larger brands benefit less than smaller brands from providing information, because they are already well known and have higher advertising goodwill and brand equity (Simon and Sullivan 1993, Dekimpe and Hanssens 1995). In other words, the incremental benefit is smaller because consumers are already more aware of the features of commonly used products.

Another theoretical underpinning to this hypothesis draws from Anderson and Renault (2009), who model advertising as revealing the direct match between consumers and horizontal product characteristics. The model predicts that smaller brands have more incentive to advertise informatively in order to differentiate themselves from larger rivals and to carve out a niche market of consumers. Conversely, larger brands might not provide information because doing so informs consumers that the brand may not be the best match for their ailments. Therefore, larger firms prefer to enhance their brands through non-informative rather than informative advertising. They might benefit from broader advertising themes, while smaller brands tend to target niche consumers with more detailed information about the product (see Iyer, Soberman, and Villas-Boas 2005). We therefore expect (as implied by Lemma 1) that

---

\(^4\) Each comparative advertisement must include at least one cue, because the comparison is made for at least one characteristic. We compare all non-comparative advertisements to all comparative advertisements, conditional on an advertisement having at least one cue.
the advertisements for larger brands will have less informative content than those for smaller brands. We formulate our third hypothesis in the following way:

**H3:** Larger brands (as measured by market share) have less information content in their advertisements than smaller brands.

**Generics**

An important characteristic of our analyzed market and of many other consumer product categories is the presence of generic substitutes, and we extend our theoretical model to allow them. If the market consists of branded and generic products, then each generic product corresponds to a particular brand, i.e. the closest substitute for that brand. Thus, informative advertisements can also increase demand for generic products as long as consumers are aware that there are a lot of shared attributes between the branded product and its generic counterpart. Thus, brands that emphasize product quality also provide free advertising for generic products. We argue that a brand’s desire to emphasize shared qualities is decreasing in the size of its generic counterpart. For example, consider an advertisement that contains partly persuasive and partly informational content. The persuasive part creates subjective differentiation and improves perceived quality only of the branded (i.e., advertised) product; the informative part increases demand for both the generic substitute and the advertised brand. Because both persuasive and informative themes have diminishing returns (for a review of advertising response functions, see Bagwell 2007), the proportion of informative to persuasive advertising should differ depending on the detrimental effect of information spillover. Thus, the larger the market share of the generic substitute, the less the brand gains from emphasizing its active ingredient–based product through informative advertising. This implies that the marginal incentive for informative content is lower for brands with generic counterparts that have large market shares. We capture these effects through an extension of the theoretical analysis described above. In the model so far, with no generics, we postulated that the increase in demand for a brand would be $D(\sum_{i=1}^{n} I_i + P(s, \varepsilon))$. Now, in the presence of a generic, suppose that the increase in demand due to advertising for the branded good is subject to some leakage to the generic products. The leakage is assumed to be increasing in the generic market share and in the information content, so that the share of customers diverted from branded products to the generic products is increasing in the generic product's market share and also in the information content of an advertisement. Furthermore, if the incremental leakage from adding an information cue is increasing in the market share of generic products, then the sizes of the information content will and the share of the

---

5 In our analyzed market, OTC analgesics, branded products tend to be equivalent in quality to their generic counterparts, due to the regulatory environment faced by this market.
generic products will be inversely related. With these stipulations in mind, we formulate our fourth hypothesis:

**H4:** Branded goods with a generic substitute with a large market share will have less informative advertising content than brands with generic substitutes which have a smaller market share.

Next, we describe the data and identification strategy that was used in the empirical application of the theoretical model presented above.

**Data and Content Analysis**

We use sales and advertising data from the OTC analgesics industry in the United States to evaluate the hypotheses constructed in the theoretical section. The OTC analgesics market covers pain relief medications with four major active chemical ingredients: aspirin, acetaminophen, ibuprofen, and naproxen sodium. The nationally advertised brands include Tylenol (acetaminophen), Advil (ibuprofen), Motrin (ibuprofen), Aleve (naproxen sodium), Bayer (aspirin or combination), and Excedrin (acetaminophen or combination).

The empirical testing of the formulated theory focuses on the OTC analgesics industry for several reasons. First, television advertising constitutes a large fraction of sales, implying that it is the most important marketing strategy used by the industry to communicate to its consumers. Second, the products sold by firms differ significantly in their characteristics, so that there is a range of meaningful information to potentially communicate (Lancaster 1971, Christou and Vettas 2008). Third, the information is concentrated in experience- and credence-based characteristics, which tend to be advertised, rather than search characteristics that consumers can learn in the store before purchase (Erdem, Keane and Sun 2008). Fourth, product differentiation—both real and spurious—is important because we emphasize the

---

6 More formally, suppose that the demand for the branded good increases

\[ D \left( \sum_{i=1}^{n} I_i + P(s, \varepsilon) \right) - L \left( s_g, \sum_{i=1}^{n} I_i \right) \]

where \( D(\sum_{i=1}^{n} I_i + P(s, \varepsilon)) \) denotes the increase in demand for both the branded product and its generic counterpart, and \( L(s_g, \sum_{i=1}^{n} I_i) \) denotes the leakage to the generic. The leakage is assumed to be increasing in the generic market share, \( s_g \), and in the information content, \( \sum_{i=1}^{n} I_i \). Note that \( n \) information cues will be preferred to \( n+1 \) if

\[ D(\sum_{i=1}^{n} I_i + P(s, \varepsilon)) + D(\sum_{i=1}^{n+1} I_i) > D(\sum_{i=1}^{n} I_i + P(s, \varepsilon)) + D(\sum_{i=1}^{n+1} I_i) \]

where we have let the argument \( \sum_{i=1}^{n} I_i \) denote the use of the first \( n \) information cues. Rewriting this condition as

\[ L(s_g, \sum_{i=1}^{n+1} I_i) - L(s_g, \sum_{i=1}^{n} I_i) > D(\sum_{i=1}^{n+1} I_i + P(s, \varepsilon)) - D(\sum_{i=1}^{n} I_i + P(s, \varepsilon)) \]

and noting that the right hand side of the equation does not depend upon \( s_g \), the lower number of cues, \( n \), is more likely to be preferred as the left hand side of the equation increases. That holds if the incremental leakage is increasing in \( s_g \). This is a natural condition given that the leakage itself is increasing in \( s_g \) (for example, if leakage were proportional to generic size times an increasing and concave function of information cues).
trade-off between persuasion and information, with the optimal mix depending upon a product’s characteristics (see Soberman 2002). For all of these reasons, the US OTC analgesics industry provides an ideal testbed for the theory. Finally, the type of cues mentioned (e.g., “strong”) are clearly identifiable, which enables us to avoid making any subjective judgments while coding the information cues.

The advertising data come from TNS Media Intelligence and cover the entire U.S. OTC analgesics product category. The data set contains video files of all advertisements, as well as monthly advertising expenditures, for each product advertised in the OTC analgesics category from 2001 to 2005. The advertising numbers also include expenditures on other media, but almost all the advertising budgets (approximately 90%) were spent on television advertising, including network and cable networks. In our analysis, we examine only the television advertising data. We watched 4503 individual commercials that were broadcast during the 2001–2005 period, 346 of which had missing video files. Each individual advertisement was usually shown multiple times.

The widely used RS method for measuring advertising information categorizes the information provided in advertisements into 14 distinct “information cues,” including price, quality, performance, components, availability, special offers, taste, nutrition, packaging, warranties, safety, independent research, company research, and new ideas. More than 60 studies have applied the RS approach to measure the information content of advertising in different media (Chou, Franke, and Wilcox 1987; Harmon, Razzouk, and Stern 1983; Stern and Resnik 1991), countries (Hong, Muderrisoglu, and Zinkhan 1987; Madden, Caballero, and Matsukubo 1986), and product categories (Stern, Krugman, and Resnik 1981). The results have varied markedly, even within the same medium, because of the lack of a multivariate statistical analysis, redundant or too broad definitions of information cues, and small sample size (Abernethy and Franke 1996). The main advantage of the RS classification system is the general nature of the information cue categories, which allow for comparison of products from multiple industries. However, this advantage could also be a disadvantage. Categorizing advertising information content into coarse categories inevitably omits some information that consumers might find important. For example, in the OTC analgesics industry, two distinct information cues (e.g., fast and strong) would be coded as one "performance" cue in the RS classification system.

Our attribute coding approach documents every attribute mentioned. For each advertisement, we recorded whether the commercial had any comparative claims and, if so, the specific claim (e.g., faster, stronger). We also noted all information cues mentioned, including the purpose of the drug (e.g., menstrual pain, arthritis, headache), drug efficiency (e.g., strength, speed), safety, and other characteristics. The type of information cues that were mentioned (e.g., “strong”) are clearly identifiable, which enables us to avoid making any subjective judgments while coding the information content. The disadvantage of this approach is that it is industry-specific. In discussing our results, we focus on the attribute coding approach, but for robustness we also present the results using the RS methodology.
Our analysis also incorporates data on the strength of pain relief, relative efficiency, and safety for each brand. We collected this information from peer-reviewed medical journals. Clinically, all four main active ingredients have varying degrees of side effects. Because individuals react to each ingredient differently, clinical pain researchers hesitate to assign superiority to any single drug. Each active ingredient has a comparative advantage. As a sample of these comparative advantages and disadvantages:

- Aspirin (brand name: Bayer) is weak in pain relief but has low, almost nonexistent cardiovascular risk.
- Naproxen sodium (Aleve) is the most potent drug but is associated with very high gastrointestinal risk.
- Acetaminophen (Tylenol and Excedrin) has low gastrointestinal risk but is weak in pain relief and has medium cardiovascular risk.
- Ibuprofen (Advil and Motrin) and naproxen sodium–based brands (Aleve) have the highest cardiovascular risk but are also the fastest in pain relief.

We quantify or rank all of the “true” characteristics that were used in advertising associated with each active ingredient as follows. First, we interpret “fast” as the time taken to achieve perceptible or meaningful pain relief (medical literature calls this “onset to perceptible pain relief”). Second, we interpret claims such as “long lasting” as the duration of meaningful pain relief. Third, we interpret claims about strength (e.g., “strong,” “stronger,” “tougher on pain”) as the maximum level of pain relief achieved; we use the number-needed-to-treat (NNT) measure to approximate analgesic efficiency claims. The NNT is a standard efficiency measure used in the pain relief evaluation literature. Table 1 compiles
spending and cue information by brand and by year under the attribute and RS coding approaches. The first two columns give each brand’s (expenditure-weighted) average number of attributes and whether it is comparative.

Figure 3. Advertised Attributes and Expenditures

During our analyzed period, 30 different product attributes were mentioned. Figure 3 represents the top 23 attributes and shows the advertising expenditures (in millions of dollars) spent on advertising those attributes during the sample period. We separate advertising expenditures by type of advertisement (comparative versus non-comparative). The attributes “fast,” “strong,” “long lasting,” and “trust and/or safety” are among the top five most heavily advertised attributes. These attributes are directly related to the inherent (exogenous) chemical characteristics of each active ingredient in each analyzed brand.

We examine the attribute usage correlations to investigate whether the coded cues represent distinct information. For example, although we code “strong” and “fast” as separate information categories, we also ensure that the coded information categories are indeed distinct information cues. Table 2 portrays the correlation matrix of cue usage and shows that the cue descriptors that we use are

---

7 The 2001 figures include only ten months of observations because we do not have the corresponding sales data for January and February 2001. The low total ad expenditures can be partially explained by this, but the average monthly expenditures should not be affected. Even when adjusted for inflation, ad expenditures tend to increase over time.

8 The remaining 7 attributes (not reported) had negligible advertising expenditures. The sum of the expenditures in Figure 3 exceeds total ad spending because many advertisements promote multiple characteristics and, for the purpose of this Figure only, we attributed total ad spending to each characteristic mentioned.
distinctive. For example, both “fast” and “strong” are often used together, but they are also used separately in more than half of the occurrences (in dollar terms). Thus, two cues may be used together frequently, but each still provides consumers with important information. “Strong” denotes how powerful the medicine is, and “fast” denotes the speed of the onset of pain relief.

Table 2. Matrix of Frequency of Attributes that are Mentioned Together

<table>
<thead>
<tr>
<th></th>
<th>Fast</th>
<th>Strong</th>
<th>Headache</th>
<th>Long lasting</th>
<th>Safe</th>
<th>Arthritis</th>
<th>Dr. recomm.</th>
<th>Liquid gels</th>
<th>Legs/muscle</th>
<th>Gentle on stomach</th>
<th>Back</th>
<th>Fewer pills</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast</td>
<td>$251.61</td>
<td>$296.52</td>
<td>$70.25</td>
<td>$32.76</td>
<td>$23.80</td>
<td>$16.69</td>
<td>$204.55</td>
<td>$101.42</td>
<td>$66.32</td>
<td>$25.65</td>
<td>0</td>
<td>$613.52</td>
<td></td>
</tr>
<tr>
<td>Strong</td>
<td>$251.61</td>
<td>$126.43</td>
<td>$103.37</td>
<td>$93.88</td>
<td>$123.06</td>
<td>$97.80</td>
<td>$140.66</td>
<td>$133.31</td>
<td>$45.46</td>
<td>$51.94</td>
<td>$53.01</td>
<td>$510.20</td>
<td></td>
</tr>
<tr>
<td>Headache</td>
<td>$296.52</td>
<td>$126.43</td>
<td>$6.22</td>
<td>$28.85</td>
<td>$11.95</td>
<td>$25.42</td>
<td>$84.88</td>
<td>$17.53</td>
<td>$24.40</td>
<td>$15.88</td>
<td>$6.22</td>
<td>$377.09</td>
<td></td>
</tr>
<tr>
<td>Long lasting</td>
<td>$70.25</td>
<td>$103.37</td>
<td>$6.22</td>
<td>$28.85</td>
<td>$11.95</td>
<td>$25.42</td>
<td>$84.88</td>
<td>$17.53</td>
<td>$24.40</td>
<td>$15.88</td>
<td>$6.22</td>
<td>$377.09</td>
<td></td>
</tr>
<tr>
<td>Safe</td>
<td>$32.76</td>
<td>$93.88</td>
<td>$68.50</td>
<td>$153.97</td>
<td>$115.84</td>
<td>$115.84</td>
<td>$125.59</td>
<td>$21.58</td>
<td>$18.38</td>
<td>$55.50</td>
<td>$39.29</td>
<td>$293.20</td>
<td></td>
</tr>
<tr>
<td>Arthritis</td>
<td>$23.80</td>
<td>$123.06</td>
<td>$153.97</td>
<td>$115.84</td>
<td>$125.59</td>
<td>$21.58</td>
<td>$18.38</td>
<td>$55.21</td>
<td>$19.22</td>
<td>$80.12</td>
<td>$80.12</td>
<td>$271.42</td>
<td></td>
</tr>
<tr>
<td>Dr. recomm.</td>
<td>$16.69</td>
<td>$97.80</td>
<td>$8.96</td>
<td>$77.64</td>
<td>$125.59</td>
<td>$50.43</td>
<td>$9.79</td>
<td>$47.43</td>
<td>$16.57</td>
<td>$70.15</td>
<td>0</td>
<td>$249.02</td>
<td></td>
</tr>
<tr>
<td>Liquid gels</td>
<td>$204.55</td>
<td>$140.66</td>
<td>$84.88</td>
<td>$14.43</td>
<td>$5.51</td>
<td>$21.58</td>
<td>0</td>
<td>$23.06</td>
<td>$44.74</td>
<td>0</td>
<td>0</td>
<td>$206.82</td>
<td></td>
</tr>
<tr>
<td>Legs/muscle</td>
<td>$101.42</td>
<td>$133.31</td>
<td>$17.53</td>
<td>$64.17</td>
<td>$29.96</td>
<td>$18.38</td>
<td>$23.88</td>
<td>$23.06</td>
<td>$23.06</td>
<td>$56.44</td>
<td>0</td>
<td>$204.79</td>
<td></td>
</tr>
<tr>
<td>Gentle on stomach</td>
<td>$66.32</td>
<td>$45.46</td>
<td>$24.40</td>
<td>$23.88</td>
<td>$23.06</td>
<td>$23.88</td>
<td>0</td>
<td>$23.06</td>
<td>$56.44</td>
<td>$7.54</td>
<td>0</td>
<td>$116.40</td>
<td></td>
</tr>
<tr>
<td>Back</td>
<td>$25.65</td>
<td>$51.94</td>
<td>$15.88</td>
<td>$44.99</td>
<td>$39.29</td>
<td>$19.22</td>
<td>$4.71</td>
<td>0</td>
<td>$56.44</td>
<td>$0.88</td>
<td>0</td>
<td>$115.98</td>
<td></td>
</tr>
<tr>
<td>Fewer pills</td>
<td>0</td>
<td>$3.01</td>
<td>$6.22</td>
<td>$100.06</td>
<td>$30.46</td>
<td>$80.12</td>
<td>$50.65</td>
<td>0</td>
<td>$7.54</td>
<td>0</td>
<td>14.83</td>
<td>$113.25</td>
<td></td>
</tr>
</tbody>
</table>

9 There are two instances of high correlation that merit comment. First, whenever "liquid gels" are mentioned, "fast" is almost always mentioned. Second, "long lasting" and "fewer pills" are used together 33% of the time. Conversely, when "fewer pills" are mentioned, long lasting was mentioned 88% of the time. In this instance, we could provide an umbrella classification that encompasses both, but the difference in results would be minor.
Identification Strategy

A brand’s decision about how much information to include in an advertisement is likely to be made simultaneously with the decision about the type of advertisement (comparative or non-comparative). Therefore, these two decisions are interdependent, in much the same way as equilibrium price and quantity are determined at the same time in a simple demand–supply model. In other words, there is some unobservable exogenous variable that explains both the information content and whether that content is a non-comparative or comparative advertisement. For example, a higher-quality brand might provide less information and have more comparative advertisements than a lower-quality brand.

The other two potentially endogenous explanatory variables that we consider are the size of the brand and the size of a brand’s generic counterpart. To see why endogeneity might be an issue, note that market size and information content are outcomes of brands’ strategic interactions, and a fully structural equilibrium model would specify three equations, one for each of the three variables (information content, market size, and the generic counterpart’s market size). Here, we estimate only one of the three equations—the one that explains information content as a function of the other two endogenous variables—but we control for the endogeneity of the other two variables. Equivalently, we could consider an unobservable variable (e.g., quality) that is correlated with the brand’s market size and information content. By omitting that variable from our regressions, we would introduce a bias into the estimation of the parameters of the model.

We use an instrumental variable approach to determine whether our concerns about the endogenous variables are empirically relevant (Villas-Boas and Winer 1999). Instruments that correlate with sales and advertising but not with unobserved quality provide information on how important the endogeneity problem is likely to be. Following the literature on the estimation of demand in differentiated product markets (e.g., Nevo 2001; Chintagunta 2001), we assume that the product characteristics space is exogenous. This is a particularly reasonable assumption for the OTC analgesics industry because pain relievers’ true characteristics are essentially fixed, determined by the chemical properties of the particular active ingredient that constitutes the drug. These active ingredients are also regulated by the FDA. Then, we consider the case in which a brand’s own characteristics enter directly into the ordered probit regression, and we construct the instrumental variables using functions (i.e., the average) of the characteristics of the brand’s competitors. More specifically, we construct both the means of the characteristics of the brand’s competitors and their minimum values. We also interact these characteristics with a dummy that is equal to 1 if a brand’s parent company owns another brand, which is the case for Tylenol and Motrin (parent company McNeil) and Aleve and Bayer (parent company Bayer). Finally, we
interact the characteristics with the 2005 year dummy to capture advertising content changes—2005 was considered one of the most turbulent years in the analgesics industry.\(^\text{10}\)

To deal with the endogenous variables in our nonlinear ordered probit model, we follow Rivers and Vuong’s (1988) proposed approach. First, we rewrite the information content as \(\alpha_t - X\beta - wy = \bar{p} - \bar{I} - \frac{I_t}{\bar{I}}\), where \(w\) is a vector of the three endogenous variables. The main identification assumption is that instruments are not correlated with the error term (i.e., \((Z'e) = 0\)). Here, \(Z\) includes all of the exogenous variables, such as brand characteristics \((X)\) and functions (here, the average) of the characteristics of the brand’s competitors. We use a two-step procedure described in Rivers and Vuong (1988). First, we run an ordinary least squares regression, \(w = Z\delta + \upsilon\), where \(\upsilon\) is not observed and is the omitted variable that generates the endogeneity problem. This first stage regression yields residuals \(\hat{\delta}\), which we include in the ordered probit in the second stage of the estimation. The estimation of this ordered probit provides consistent estimates of all of the parameters. Here, we are not interested in the magnitude of the parameters of the information content relationship (i.e., \(\beta\)) but rather in the marginal effects of a change in the endogenous variables, which we can consistently estimate using Blundell and Powell’s (2003) approach.

### Empirical Findings

Table 3 reports the findings of the various specifications estimated with our data. The first five columns present the results from estimating our model with the attribute coding. For robustness, in the last column we also include results estimated using the RS information coding.

**H1: Vertical Characteristics**

We test H1, i.e. that brands with higher vertical quality transmit more information, by associating the number of cues (our dependent variable) with the values of the exogenous medical characteristics of the active ingredients. The first column of Table 3 includes NNT, relative speed, and measures of cardiovascular and gastrointestinal risk as such explanatory variables.

Keeping all else constant, we find that brands with inherently stronger pain relief have advertisements with more information content. Recall that for NNT, a higher number means worse performance and a less effective drug. Thus, the negative coefficient on NNT is consistent with strong

---

\(^{10}\) The growth of information content during 2005 is most likely due to the FDA’s announcement of the results of a clinical study, at the end of 2004, which indicated that patients taking naproxen sodium (Aleve) may be at an increased risk heart attack or stroke (the withdrawal of Vioxx was also associated with this clinical study). By the end of January 2005, sales of Aleve plummeted by more than 50%, suffering the largest decline in brand history (for more details, see ARF Ogilvy Awards 2007).
and efficient drugs having more information in their advertisements. We find a similar pattern for brands that have lower gastrointestinal and cardiovascular risks: Their advertisements also tend to be more informative. Finally, brands that offer faster relief also have more information content.

Overall, the first column of Table 3 suggests that there is a positive relationship between the amount of information provided and the strength of a painkiller along one of the four dimensions identified by the exogenous medical characteristics of the active ingredient. This finding provides empirical support for H1.

As we discussed in our theoretical section, the cutoffs estimated in these regressions have a clear structural interpretation. In particular, we showed that \( a_i + 1 - a_i = \frac{I_i - I_{i+1}}{\bar{s}} \). In our analysis \( \bar{s} \) is unknown, so assume for the sake of interpretation that \( \bar{s} = 1 \). Then, for each \( i \) we can compute the difference \( I_i - I_{i+1} \), which is the incremental benefit of the \( i \)th information cue \( i \). Consider the case of \( i=5 \). Then, \( I_5 - I_6 = 1.002 \). We can then find the point \((6,1.002)\) in Figure 4 to represent the empirical analogue of the theoretical incremental benefit corresponding to six information cues that we had derived in Panel B of Figure 2. We can replicate this exercise for the case of \( i=4 \), for which the scatter point will correspond to \((5,1.939)\), which is equal to \( 1.002 + 0.937 \). By repeating this exercise for all \( i = 1, \ldots, 6 \), we derive the scatter plot in Figure 4. What we see in Figure 4 is remarkably similar to what we derive theoretically in Panel B of Figure 2, which provides empirical support for our theoretical model: there is a fundamental trade-off between information and persuasion content that firms face when preparing an advertisement. This relationship, depicting the diminishing marginal returns to information, holds in every specification that we estimate and looks very similar to the one illustrated in Figure 4.

**H3 and H4: Brand Market Size and Generic Counterpart Market Size**

To test H3 and H4, we include the variables that measure brand size, brand size squared (to capture a possible nonlinear relationship), and the size of the brand’s generic counterpart. Table 3 presents the results of the ordered probit with standardized sales, squared standardized sales, and standardized generic sales. In this specification, we conclude that the largest brands transmit less information than the rest of the brands. There are very few observations for medium-sized brands, therefore it is difficult to conclude whether the negative size-information content relationship holds over the entire range of brand sizes (the positive coefficient on the brand size variable and the negative coefficient on the quadratic term suggest that the relationship resembles a weak inverted U-shape).

\[ \text{Instead of normalizing } \bar{s} = 1 \text{, one could also report } \frac{a_i + 1 - a_i}{a_0 - a_1} \text{ but that would only rescale Figure 3. The normalization } \bar{s} = 1 \text{ gives a more intuitive interpretation for Figure 3.} \]
Table 3. Determinants of Information Disclosure

<table>
<thead>
<tr>
<th></th>
<th>Attribute Coding</th>
<th>R-S Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Comparative?</td>
<td>0.559***</td>
<td>2.115***</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.099)</td>
</tr>
<tr>
<td>Standardized Sales</td>
<td>0.445***</td>
<td>0.268***</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.089)</td>
</tr>
<tr>
<td>Standardized Sales Squared</td>
<td>-0.204***</td>
<td>-0.160***</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Standardized Generic Sales</td>
<td>-0.136***</td>
<td>-1.571***</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.084)</td>
</tr>
<tr>
<td>Standardized NNT</td>
<td>-0.724***</td>
<td>-0.602***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Relative Speed</td>
<td>0.336***</td>
<td>0.281***</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Standardized GI Risk</td>
<td>-0.133***</td>
<td>-0.132***</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>Standardized CV Risk</td>
<td>-0.151***</td>
<td>-0.185***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Residuals-Comparative</td>
<td>-1.652***</td>
<td>-1.103***</td>
</tr>
<tr>
<td></td>
<td>(0.102)</td>
<td>(0.178)</td>
</tr>
<tr>
<td>Residuals-Sales</td>
<td>0.174</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.117)</td>
<td></td>
</tr>
<tr>
<td>Residuals-Generic Sales</td>
<td>1.844***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.112)</td>
<td></td>
</tr>
</tbody>
</table>

Cutoff (0->1 Cues)            -3.085***
(0.194)

Cutoff (1->2 Cues)            -1.041***
(0.090)

Cutoff (2->3 Cues)            0.291***
(0.089)

Cutoff (3->4 Cues)            1.095***
(0.090)

Cutoff (4->5 Cues)            2.032***
(0.091)

Cutoff (5->6 Cues)            3.034***
(0.097)

# of Obs.                     9,739 9,708 9,708 9,708 9,708 9,708
Log-Likelihood               -13847.5 -13527.6 -13394.9 -13378.8 -13159.5 -9476

Note: *** p<0.01, ** p<0.05, * p<0.1; Bootstrapped standard errors reported in Columns 2, 5 and 6. Endogenous variables: Comparative?, Standardized Sales, Standardized Generic Sales. Instruments: (1) averages of Standardized GI Risk, Standardized CV Risk, Standardized NNT, Relative Speed; (2) Minimums of Standardized GI Risk, Standardized CV Risk (3) Interactions between characteristics and year 2005 dummy; (4) Interactions between characteristics and a dummy indicating whether a brand has a parent company that owns competing brand in the category. The first stage R^2 for endogenous variables is the following: Comparative? - R^2 = .28, Standardized Sales - R^2 = .95, Standardized Generic Sales - R^2 = .93)
The evidence is also consistent with a relatively large spillover effect from informative advertising. We find that branded firms include less information content when the size of their generic competitors is large because the parameter of standardized generic sales is negative.

We also find that the coefficient estimate for the comparative advertisement dummy is close to the one in the second column of Table 3. Therefore, we conclude that comparative advertising decisions are not collinear with brand sales and generic sales. This result is noteworthy because it suggests that the comparative advertisement decision might not depend significantly on the market share of the attacking brand but rather on the market share of the target brand (i.e., attack larger brands). Anderson et al. (2011) investigate the relationship among the attacker’s market share, the attacked brand’s market share, and the amount of their comparative advertising. They show that the decision to use comparative advertising is due to the market share of the attacked brand and the interaction between the shares of the attacked brand and those of the attacking brand.

In the fifth column of Table 3 we follow an instrumental variable approach to estimate the effect of comparative advertisements, sales, and generic sales on information content. The estimated coefficient of comparative advertisements is large, although not as large as in the third column. The coefficient on standardized sales is similar to the coefficient under exogenous treatment. However, the marginal impact of the generic counterpart’s market share is significantly greater under endogenous treatment, but continues to be negatively related to information content.

This result implies that when endogeneity of the generic counterpart market share is not accounted for, its negative relationship with advertising content is significantly underestimated. In line with the exogenous treatment in the fourth column, we observe the same relationship between the brand
size and information content. Thus, these results indicate that the largest brands include less information content.

We also observe that the coefficient estimates of the control functions for generic sales (1.844) and for the decision to have comparative advertising (-1.103) are both statistically significant, proving that both variables (generic sales and comparative advertisements) are endogenous. The estimated coefficient of the control function for brand sales (0.174) is only statistically significant at the 13.7% level. This finding is probably due to our instrumental variables having just enough identification power to identify two endogenous variables.

Figure 5. Marginal effects on Information Disclosure by Brand Size
To understand the economic importance of the results regarding brand market size, we constructed figures that associate a brand’s probability of choosing a given amount of information (e.g., one cue) with the distribution of brand size (at the 10th, 25th, 50th, 75th, and 90th percentiles of market size distribution). For example, Figure 5 shows that the likelihood of an advertisement including only one cue increases sharply with size. In contrast, the probability of observing an advertisement with four or more cues decreases with size. For example, a move from the 25th percentile in the size distribution to the 75th percentile increases the probability of providing only one cue by approximately 25%.

Figure 6. H4: Marginal effects on Information Disclosure by Generic Counterpart Size

Similarly, Figure 6 represents the changes in the likelihood that an advertisement includes a certain number of information cues depending on the size of the generic counterpart market share (again,
evaluated at the 10th, 25th, 50th, 75th, and 90th percentiles of generic counterpart market size distribution). For example, the likelihood that a product with a large generic counterpart will include one information cue is 80% higher than the likelihood that a product with a small generic counterpart will include one information cue. The converse is true for three or more information cues: brands with smaller generic counterparts are significantly more likely to include more information in their advertisements.

In summary, we find weak support for $H_3$ and strong support for $H_4$.

**RS Approach**

The last column of Table 3 presents the results of the estimation using the RS information category coding approach. We implement this step to evaluate the extent to which our results are robust to alternative information content definitions. We find that all of our hypotheses with the exception of $H_1$ are also supported with the RS coding approach. All of the key variables (except for vertical quality) have the same signs as those with the attribute coding approach. The magnitudes are smaller because the left hand side variable (information content) is smaller under the RS coding approach. The vertical characteristics parameter estimates under the RS information coding approach are different for strength (NNT) and speed. As we explained previously with regard to this mismatch, under the RS approach both characteristics fall under one information category and therefore understate the information content of advertising.

**Final Remarks**

This research provides several contributions to the literature on the information content of advertising. First, we develop a theoretical model that describes the tradeoff between persuasive and informative content within an advertisement. We use the theoretical model to predict the relationship between the informative content and various market performance indicators. The empirical component of the paper applies the theory to the OTC analgesics advertising content data. The analysis gives rise to a deeper and broader description of information content, which combines advertising content knowledge with industry structure and advertising spending data.

In line with our theoretical results of a trade-off between informative and persuasive advertising, we show empirical evidence that supports our hypotheses: we find that brands with higher vertical quality disclose more information, and advertisements for brands with a high market share provide less information. These two results are not contradictory. The endogeneity of brand size underscores the importance of correcting for it. Otherwise, it might be assumed that larger brands are fundamentally of higher quality than smaller brands and, therefore, that their advertisements should have more information content. Correcting for endogeneity, we also show that more competition from generics gives rise to less
information transmission by branded products, which is in line with the view that there are significant spillovers to informative advertising. Finally, we quantify the extent to which comparative advertisements have significantly more information content than non-comparative ones, and find that this effect is much larger than would be predicted without correcting for endogeneity.

From a methodological standpoint, we describe a method that is appropriate for dealing with information content with multiple explanatory variables, and show how the analysis can be corrected for endogeneity and how this alters the results. Our empirical analysis is restrictive in several aspects, and therefore suggests extensions that constitute themes for further research into information content. First, further research could use a sub-classification of cues (e.g., into vertical cues such as “fast” that all consumers would appreciate, or horizontal cues such as “headache” or “menstrual pain” that only some consumers desire) to explore the differential content of various cue types. Second, an information cue can be deployed only if a product has the attribute communicated in the cue, and can be used comparatively only if a product has an advantage over another product. Thus, further research might examine the amount of information advertised as a function of the total number of cues that could feasibly be advertised. Likewise, do products use comparative advertisements more often against similar or dissimilar products? Third, we do not examine product advertising campaigns, in which advertisements address a subset of themes over a limited horizon. Fourth, we code only the objective content of advertisements as quantified by their reference of specific characteristics and competitors. We recognize that advertising may persuade through channels other than pure information, thereby leading consumers to act on emotional factors. We have not attempted to code such effects, but our theoretical model allows for the existence of such effects. The primary purpose of the empirical component of this paper was to measure the objective content of advertising along the lines of traditional content analysis, and incorporating the subjective side would be an important aspect to explore in future research. Finally, we do not address whether the market provision of information is optimal, how valuable the information is to consumers (Ippolito and Pappalardo 2002; Pappalardo and Ringold 2000) or how government policy towards advertising content might affect promotion effectiveness (Goldfarb and Tucker 2011). The purpose here was to document and, using our theoretical model, to rationalize empirical regularities present in the data and provide measures of the fundamental variables that can be used to answer such questions.

Our theoretical model and content analysis methodology can readily be applied to other product categories and industries. Extensions of the proposed methodology could be applied to answer questions such as: How does advertising information content differ for experience, search, and credence products? Does the relationship between brand size and information hold in more broad contexts? Do new products provide more information? These questions should hold both empirical and theoretical interest.
REFERENCES


Appendix. Content Analysis

Expenditure-Weighing

Many traditional content studies do not include data on what brands paid to screen, air, or publish advertisements, so only information cues per advertisement analyzed are reported. As we discussed previously, our data include the complete set of advertisements run over the sample period and the expenditure and airing frequency data. These data provide different ways to count the observations. Figure A1 represents the histograms of information cue distributions under varying approaches. The first columns in Figure A1 weight each advertisement occurrence by its advertising expenditures. The "unweighted" data in Figure A1 use each separate advertisement within any given month as an observation. In this case, multiple airings of the same advertisement within the same month are ignored. This is in line with most of the traditional content analysis studies: Multiple copies of the same advertisement were typically not counted as different observations.

Figure A1. Distribution of Number of Cues

The distribution of information cues across the different measures is similar, which indicates that a lack of data on the amount spent on airing the advertisements does not distort the results for our particular sample. The distribution of cues for the unweighted data stochastically first-order dominates the distribution for the weighted data -- there is a greater fraction of ads with zero cues, a greater fraction with one or fewer, etc. This implies that more money tends to be spent on running ads with more cues in them. In other words, if ads were all aired the same number of times, those with more cues are being aired to more expensive (i.e., typically larger) audiences.