August 2022

Three Essays on Online Customer Reviews in Business and Healthcare

Cong Zhang
University of Wisconsin-Milwaukee

THREE ESSAYS ON ONLINE CUSTOMER REVIEWS IN BUSINESS AND HEALTHCARE

by

Cong Zhang

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Management Science

at

The University of Wisconsin-Milwaukee

August 2022

ABSTRACT

THREE ESSAYS ON ONLINE CUSTOMER REVIEWS IN BUSINESS AND HEALTHCARE

by

Cong Zhang

The University of Wisconsin-Milwaukee, 2022
Under the Supervision of Professor Atish Sinha

As e-commerce platforms keep growing in popularity, online customer reviews, which represent users' evaluation of products or services, have become a crucial information source in consumer decision-making. Online reviews have proven to have a major impact on various critical aspects of a business, such as reputation, sales, and product returns. The goal of the three-essay dissertation is to investigate the influential antecedents and consequents of online customer reviews in business and healthcare.

Essay I explores the effects of product exposure time on review content and review helpfulness. We find that the descriptions of utilitarian attributes in a review increase with product exposure and mediate the relationship between product exposure and review helpfulness. To test the effects of product exposure, we first extract latent topics from review content and then identify utilitarian topics from them. Next, we build a regression model to test the utilitarian information's relationship with product exposure. The results support our central thesis that product exposure has a significant positive influence on review helpfulness, and this relationship is mediated by the utilitarian information in a review. We also find that users who have prior knowledge of the domain the product belongs to do not need long exposure times to write helpful reviews. Our findings demonstrate the need to account for product exposure and domain knowledge when examining online review helpfulness. The finding that early reviews tend to be

less helpful because they contain less utilitarian product information has important implications, both for research and for practice.

Essay II focuses on investigating how medical performance factors affect hospitals' online ratings. We find that readmission, mortality, safety of care, and time in emergency department significantly influence a hospital's online reputation. We also extract three influential review content factors: reviewer medical knowledge, medical quality evaluations, and service quality evaluations. This is the first study to investigate the effect of a set of representative hospital medical performance factors on online ratings. Furthermore, it is the first attempt at examining the roles of reviewer medical knowledge and different types of experiential quality evaluations in the online healthcare review domain. We also find a significant influence of CMS overall quality star rating on a hospital's online reputation. The findings provide valuable inputs into a hospital's marketing strategies and have important managerial implications for providers, patients, and online platforms.

In Essay III, we propose a hybrid aspect-based sentiment analysis (ABSA) framework that mines patients' online evaluations of a hospital from different aspects. We then integrate the extracted average sentiment polarities into regression models, where the numeric online rating is the dependent variable. The results show that including the aspect categories' polarities dramatically increases the models' fit. The standardized coefficients reveal that "Staff," "Nurse," and "Doctor" are the three most influential aspects of hospitals' online review ratings. Our results prove the necessity of adopting ABSA in the online healthcare review domain. They also have practical implications for patients, healthcare providers, and online review platforms.

To

My mom,

my dad,

and all my dear friends

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to sincerely thank my professor, research advisor, and dissertation committee chair, Dr. Atish Sinha. He has supported and encouraged me during the whole process. He has brought valuable opportunities for me in teaching and research. He has made a remarkable difference in my life as he made me see a possible future in academia. I could not be more grateful to have such an extraordinary researcher and educator as my advisor. I would also like to extend my appreciation to my committee members: Dr. Arun Sen, Dr. Mark Srite, and Dr. Yang Wang. I appreciate their essential feedback and insights during my dissertation. I could never complete this dissertation without their help. I thank Dr. Huimin Zhao for recognizing my potential and bringing me to the program. I also appreciate his valuable advice as my academic advisor. Thanks to Dr. Derek Nazareth for patiently guiding me with my first academic project.

I want to express my genuine appreciation to Mr. and Mrs. Lubar for their generosity in supporting my research financially. I would like to extend my thanks to all the faculty and staff at Lubar College of Business for supporting me. Especially, I would like to thank Dr. Patrick Doran and Dr. Odai Khasawneh for their guidance in my teaching, Dawn Koerten and James Hardy for helping me with the Ph.D. program affairs, and Dr. Ross Hightower, Sandy Mollberg, and Andrea Zweifel for helping me with my work for the Center for Technology Innovation.

I wish to sincerely thank my parents for their unconditional love and support. They have raised me up to be a good person and supported me in pursuing my goals. Thanks to them for always being there for me. Last but not the least, I thank myself for the courage to overcome all the barriers and never give up.

Essay I:

The Influence of the Interplay of Product Exposure Time, Utilitarian Information, and Domain Knowledge on Review Helpfulness

Abstract

This study explores the effects of product exposure time on review content and review helpfulness. We find that the descriptions of utilitarian attributes in a review increase with product exposure and mediate the relationship between product exposure and review helpfulness. To test the effects of product exposure, we first extract latent topics from review content and then identify the utilitarian topics from them. Next, we build a regression model to test the utilitarian information's relationship with product exposure. The results support our central thesis that product exposure has a significant positive influence on review helpfulness, and this relationship is mediated by the utilitarian information in a review. We also find that users who have prior knowledge of the domain the product belongs to do not need long exposure times to write helpful reviews. Our findings demonstrate the need to account for product exposure and domain knowledge when examining online review helpfulness. The finding that early reviews tend to be less helpful because they contain less utilitarian product information has important implications, both for research and for practice.

Keywords: online customer reviews, product exposure time, domain knowledge, review topics, review helpfulness, mediation relationship

Introduction

As e-commerce platforms keep growing in popularity, online customer reviews, which represent users' evaluation of products or services, have become a crucial information source in the consumer decision-making process. Online reviews, which are the most common way to express customers' product evaluations, have proven to be a critical factor influencing business reputation, sales, and several other key aspects of a business (Chevalier & Mayzlin, 2006; Clemons et al., 2006; Luo & Zhang, 2013). Because of the importance of online reviews to customers, sellers, and online platforms, almost all the major online review platforms currently have mechanisms to remind purchasers, via email/text, to post their product evaluations. For example, Amazon.com has a system to send reminders to customers within 30 days of delivery[1]. An important question that has not been addressed so far is how long an online platform or seller should wait before reminding its customers to post reviews. Many sellers encourage their customers to post reviews as early as possible. However, they ignore a critical aspect of online reviews – their *helpfulness*. A hasty product evaluation may not provide future customers with relevant and comprehensive information because the reviewer does not have enough time to experience the product. Albeit logical, whether the exposure time is positively correlated with the review helpfulness is unclear. A major goal of this study is to answer this research question.

Product evaluations come mainly from users' memories and their experiences while interacting with the products (Vermeeren et al., 2010). A user needs to spend a significant amount of time experiencing the product. Accordingly, we raise the following question: Does a user's time of exposure to a product affect the helpfulness of her online reviews? For instance, a

[1] https://sellercentral.amazon.com/gp/help/external/help.html?itemID=G1701

user may choose to write a review one week after her purchase. She can also choose to write a review a month later. In such a scenario, we want to examine whether a review posted a month later would be voted as more helpful than a review posted shortly after the consumption. The results could have major business implications if we find evidence to support the effect of time on review helpfulness. Sellers and online review platforms can adjust their strategies by expediting or delaying customer reminders to receive more helpful reviews. Moreover, if the product exposure time is an effective indicator of review helpfulness, platforms will have the ability to identify potentially helpful reviews within a short time after posting and prioritize those reviews on their web pages. We also raise the following question: Does a user who has prior knowledge of the domain the product belongs to need the same level of exposure as a first-time buyer to post a helpful review?

Most e-commerce platforms allow consumers to vote for reviews that they consider helpful. Because helpful reviews are diagnostic and weigh more heavily in consumer purchase decisions (Yin et al., 2016), investigating review helpfulness is an important research question (Mudambi & Schuff, 2010). Therefore, many online review helpfulness studies have emerged (Ghose & Ipeirotis, 2010; Yin et al., 2020). Researchers have discovered various influential factors such as review rating, review length, and review subjectivity. However, temporal factors have surprisingly attracted limited attention. Our focus in this essay is on the effect of product exposure on online review helpfulness. The focal temporal factor, *product exposure*, is measured as the time difference between the product shipping date and review date. Our preliminary analysis indicates that product exposure correlates strongly with self-reported product familiarity levels, suggesting that this factor can be used as a proxy for customers' product experience. The results of our main estimations indicate that product exposure is positively associated with review helpfulness. It means that the longer the users are exposed to the products they purchase, the more helpful their reviews will be.

We next delve into the question: Why do reviews with longer product exposure tend to attract more helpfulness votes? Although no study in the online market has addressed this question, there is a stream of research that analyzes user experience changes over time in offline settings (Karapanos et al., 2010; Karapanos et al., 2009; Kujala et al., 2011). These studies have found that as users' interaction time with products accumulates, their product evaluations increasingly focus on product functionality – a utilitarian product characteristic. Inspired by such findings, we dive deeper to examine the impact of product exposure on utilitarian information in a review.

To address the proposed research questions, we first extract latent topics from review contents using the Topic Modeling approach and name them based on our interpretations of the keywords and representative reviews. Following previous studies in product feature classification (Hirschman & Holbrook, 1982), we identify three groups: *utilitarian information*, *hedonic information*, and *other information*. We then investigate the relationship between utilitarian information and product exposure. The results suggest that reviewers with longer product exposures tend to write online reviews that contain more utilitarian information. Next, we analyze the impact of utilitarian information on review helpfulness. Our analyses suggest that utilitarian information enhances review helpfulness. Overall, we identify a significant mediation relationship, where the effect of product exposure on review helpfulness is mediated by utilitarian information.

Moreover, we examine if a user's knowledge of the domain the product belongs to moderates the effect of product exposure on review helpfulness. The results suggest a negative moderating effect, implying that the positive effect of product exposure is mitigated by domain knowledge. Finally, we conduct several robustness checks to validate our findings.

Our study offers important implications, both with respect to theory and practice. Online review helpfulness has attracted a lot of attention recently, but temporal factors have been largely ignored This study is the first attempt to investigate the interplay of product exposure, utilitarian information, and domain knowledge and its impact on review helpfulness. Our results help better understand the formation of review content and review helpfulness. Furthermore, our findings may help e-commerce participants improve their reminder strategies, review guidelines, and marketing tactics. For example, sellers can postpone sending reminders since longer product exposure times lead to more helpful reviews. Moreover, in their review guidelines, they can encourage reviewers to write more about their evaluations of a product in terms of its utilitarian aspects to increase the likelihood of receiving more helpful reviews. Companies can also focus on illustrating their products' functional performance in their marketing channels.

The rest of the essay is arranged as follows. We first review the relevant literature related to product review helpfulness, consumer knowledge, product attributes, and user experience. Next, we develop the theoretical framework for our study and state the research hypotheses. We then describe our data and method, followed by the empirical results. Finally, we discuss the business implications of our empirical findings and conclude the essay by pointing out limitations and future directions.

Literature Review

Online Review Helpfulness

Most online product review platforms adopt a voting system that allows the customers to indicate whether a review is helpful. A helpful review is defined as a peer-generated product evaluation that facilitates the consumer’s purchase decision process (Moore, 2015; Yin et al., 2014). Several studies have examined the factors that affect review helpfulness. These factors can be grouped into review features such as review sentiment, rating, and length of the review (Agnihotri & Bhattacharya, 2016; Baek et al., 2012; Cao et al., 2011; Forman et al., 2008; Huang et al., 2018), review content variables such as subjectivity and readability (Ngo-Ye & Sinha, 2014; Weathers et al., 2015), and reviewer characteristics such as reviewer reputation and reviewer image (Chua & Banerjee, 2015; Ma et al., 2013). Mudambi and Schuff (2010) found that review extremity, review depth, and product type affect a review’s perceived helpfulness. Another study revealed that review subjectivity, readability, and spelling errors could significantly affect product sales and perceived usefulness (Ghose & Ipeirotis, 2010). Yin et al. (2014) examined the effects of two negative emotions – anxiety and anger – embedded in an online review on its perceived helpfulness. The results showed that anxiety-embedded reviews are perceived as more helpful than anger-embedded reviews, and perceived cognitive effort mediates this effect. In a later study, they further investigated the effect of anger on online reviews and found that although anger decreases review helpfulness, it has a greater impact on readers’ attitudes and choices (Yin et al., 2020).

Another popular research stream regarding product reviews is the use of topic modeling to extract latent topics from review texts (Buschken & Allenby, 2016; Korfiatis et al., 2012; Puranam et al., 2017). To understand factors that influence online food purchases, Heng et al. (2018) extracted four latent topics from product reviews and found that three of them have a significant impact. Specifically, they found that readers perceive objective information, such as physical and flavor features, as being more helpful than subjective information (e.g., personal opinions).

Although previous studies have identified various influential factors for review helpfulness, temporal factors have drawn little attention. The prior studies that investigate temporal factors commonly focus on the temporal distance of online reviews. Huang et al. (2018) find temporal cues significantly moderate the effect of product review content on review helpfulness. They operationalize temporal cues as timestamps of the date on which the product reviews were posted. Similarly, Lu et al. (2018) find that temporal factors significantly moderate the effects of static drivers (valence, length, expertise, and trustworthiness) on review helpfulness. They introduce two temporal factors, namely post lifespan and post timing. The post lifespan is measured as the time difference between the review posting date and the data collection date. The post timing is measured as the time difference between the product release date and the review posting date. However, this study focuses on how long a user has exposed to her product and its impact on review helpfulness.

To the best of our knowledge, none of the existing studies has explored the effect of product exposure – a temporal factor – on review helpfulness. Users' product experience time has been a crucial factor in the formation process of product evaluations (Karapanos et al., 2010; Kujala et al., 2011). Thus, understanding the role of user experience time in online review formation and helpfulness is essential for online businesses. To fill this gap in the literature, we propose a proxy, *product exposure*, for the actual user experience time.

Product Familiarity, Consumer Expertise

Product familiarity is defined as the number of product-related experiences that have been accumulated over time by the consumer (Alba & Hutchinson, 1987). Consumer expertise or knowledge refers to the ability to perform product-related tasks mainly gained through product experiences (Alba & Hutchinson, 1987). Many studies investigated their effects on various constructs, especially product evaluation (Bettman & Park, 1980; Cordell, 1997; Heimbach et al., 1989; Johnson & Russo, 1984; Maheswaran et al., 1996; Park & Lessig, 1981; Raju & Reilly, 1980). Raju (1977) unveiled that product familiarity can positively affect brand selection confidence. Also, higher product familiarity creates more discriminative product evaluations. Maheswaran (1994) explored the moderating effect of consumer expertise on the relationship between the original country of a product and product evaluation. Cordell (1997) found that consumer expertise moderates the evaluation of extrinsic product cues.

Consumer expertise and knowledge have also been widely investigated in online business domains (Cheung et al., 2012; Kim et al., 2011; Zou et al., 2011). In a study examining the effects of consumer knowledge on online customer reviews, researchers showed that the effect of cognitive fit (between type of review and level of consumer expertise) on purchase intention is stronger for experts than novices (Park & Kim, 2008). Ketelaar et al. (2015) examined the effects of the expertise of receivers of online reviews. The results suggest a moderating role of receiver expertise for both the influence and the weight of review valence effects.

Although consumer expertise is found to be a significant moderator for many relationships, to our best knowledge, no study has explored the possible moderating effect of consumer expertise on the relationship between product exposure and review helpfulness. By filling this gap, we test the moderating role of domain knowledge in the relationship between product exposure and review helpfulness.

Product Attribute Groups and User Experience

To better understand the effects of different product attributes, prior studies proposed and justified various product attribute groups based on their common characteristics (Brown, 2003; McGinnis & Ullman, 1992). Among all the different classifications, utilitarian versus hedonic is one of the most accepted. It has been developed and adopted for many years, starting from a series of articles by Hirschman and Holbrook (Hirschman & Holbrook, 1982; Holbrook & Hirschman, 1982). Generally speaking, utilitarian attributes are primarily instrumental and functional, whereas hedonic attributes are more related to subjective sentiments (Hirschman & Holbrook, 1982; Strahilevitz & Myers, 1998). Later on, researchers in the marketing area investigated different types of product attributes on various factors, including consumer choice (Dhar & Wertenbroch, 2000) and consumer satisfaction (Botti & McGill, 2011).

Another popular classification of product attributes is pragmatic versus hedonic, first proposed by a series of studies by Hassenzahl (2001) and primarily accepted in the Human-Computer Interaction (HCI) field. The definitions of the two classifications (i.e., utilitarian versus hedonic and pragmatic versus hedonic) are very close[2]. Pragmatic attributes are goal-oriented and are closely related to utility and usability. Utility means the ability for a product to provide the relevant functionality for performing tasks, and usability refers to whether users can access the functionality easily and efficiently (Hassenzahl, 2001). On the other hand, hedonic attributes are non-instrumental. They are self-oriented and reflect the extent to which a product is enjoyable to use (Hassenzahl et al., 2000).

Following the studies by Hassenzahl, a new research stream, user experience (UX), emerged in the Human-Computer Interaction (HCI) field. The studies on UX focus on the changes in user experience towards different types of product attributes over time (Hassenzahl, 2008; Law et al., 2009). To study how user experience develops over time, Karapanos et al.

[2] We consider the term "pragmatic attributes" and "utilitarian attributes" as the same thing. Because "utilitarian attributes" is widely adopted in the business domain, we mainly use that term henceforth.

(2009) designed an experiment to track and record participants' uses and evaluations over four weeks. The results suggest that while early experiences are mainly related to the hedonic aspects of product use, prolonged experiences become increasingly tied to aspects reflecting how the product becomes meaningful in one's life. Followed by their initial framework, Karapanos et al. (2010) presented an innovative approach that captures user experience changes over time. Consistent with their prior studies, they found that the dominance of learnability and stimulation experiences decreased over time; meanwhile, usefulness and long-term usability started to gain importance.

Prior studies that examined the temporality effect on user experience development were conducted in offline settings using traditional longitudinal and retrospective approaches, such as questionnaires (Olsson & Salo, 2012), interviews (Woo & Lim, 2002), and activity logging (Staiano et al., 2012). However, to the best of our knowledge, no study has investigated how online review contents change over time.

Other than the temporality effect on different product attributes, previous studies have also examined their impacts on perceived review helpfulness (Ham et al., 2019; Hazari et al., 2017; Kronrod & Danziger, 2013). For example, Moore (2015) studied the moderating effect of a product attribute type (utilitarian vs. hedonic) on the relationship between explanation type (action vs. reaction) and review helpfulness. The results indicate that review readers regard explained actions as more helpful than explained reactions for utilitarian products and explained reactions as more helpful than explained actions for hedonic products. Yin et al. (2017) explored how expressed emotional arousal in an online review affects its review helpfulness. The results showed an inverted U-shaped relationship, and this nonlinear effect is partly mediated by perceived effort. Moreover, this nonlinear effect is stronger for utilitarian products than hedonic products.

As far as we know, no study has investigated the mediation relationship between product exposure time and online review helpfulness. Note that, instead of identifying an entire product or a review as utilitarian or hedonic, we assume a product can have both utilitarian and hedonic attributes. As a result, its review could contain evaluations for both types of product attributes but differ only in terms of their contributions. Therefore, we extract and aggregate descriptions of utilitarian and hedonic attributes from online reviews. Next, we include this information in our models to explore its role in the relationship between product exposure and review helpfulness. Also, to the best of our knowledge, this is the first study to examine if the chunks that users with prior domain knowledge possess facilitate evaluations of product functionality and help improve review helpfulness.

Table 1.1 summarizes relevant studies in the area. Among the online review helpfulness studies shown in the table, none of them examines the temporal effects of product exposure, a temporal factor, and utilitarian product information in a review.

Table 1. 1 Summary of Literature.

Study	Outcome	Key factors	Temporal factor	Utilitarian Information
Forman et al., 2008	Review helpfulness rate and product sales	Reviewer disclosure of identity descriptive information, Shared geographical location, Review equivocality	None	No
Ghose & Ipeirotis, 2010	Review helpfulness and product sales	Review subjectivity, Review readability, Reviewer disclosure, Reviewer history	None	No
Mudambi & Schuff, 2010	Review helpfulness	Review extremity, review depth, product type	None	No
Weathers et al., 2015	Review helpfulness	Review diagnosticity, Review credibility, Product types	None	No
Yin et al., 2014	Review helpfulness	Anxiety, Anger	None	No
Yin et al., 2016	Review helpfulness	Review rating, Rating deviation, Dispersion of ratings	None	No
Yin et al., 2017	Review helpfulness	Emotional arousal, Perceived effort	None	No
Yin et al., 2020	Review helpfulness and attitude	Anger	None	No
This study	Review helpfulness	Product exposure time, Utilitarian information, Domain knowledge	Product exposure time	Yes

Theoretical Framework

In this section, we develop the theoretical framework for our study and state the research hypotheses. Exposure is mainly measured as the repetitions of a stimulus (Zajonc, 1968; Zajonc et al., 1974). Stang (1973, 1975) contends that repeated exposure gives more chance to learn about the stimulus. In the business area, product familiarity is closely related to product usage experience or frequency of use (Anderson et al., 1979; Jacoby et al., 1978). A longer product exposure allows customers to interact more with their products and accumulate more product-related experiences, thus increasing product familiarity.

As users' interaction time with products accumulates, their product evaluations focus more on product functionality (Karapanos et al., 2010; Karapanos et al., 2009; Kujala et al., 2011), a utilitarian product characteristic. When a reviewer gets exposed to a product for longer periods of time, she develops a more comprehensive understanding of the functions it offers and how well it delivers those functions. Hence, her reviews will tend to incorporate descriptions of the product's utilitarian attributes in general and its functional performance in particular.

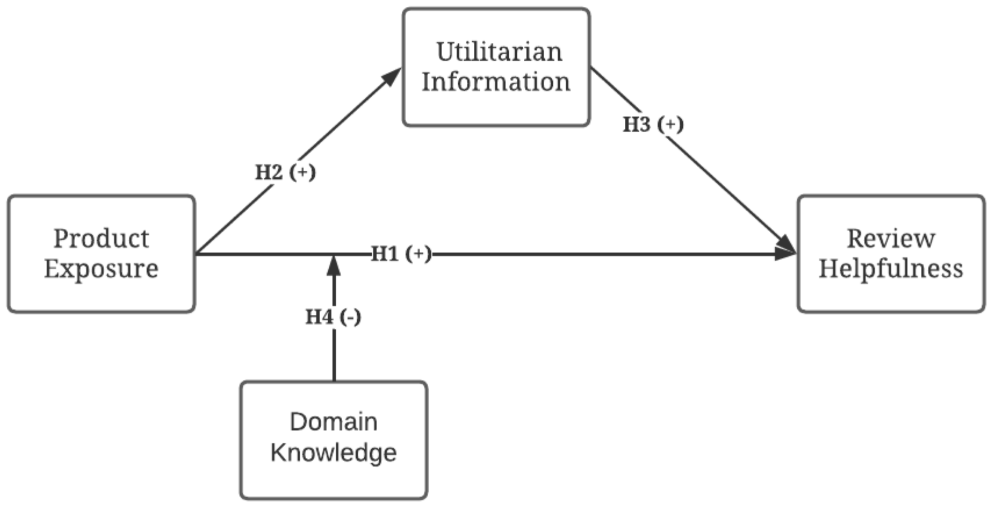
Those users who purchased a product from the same product group in the past would not need the same exposure time to provide such functional information in their reviews. Past studies suggest that users who have prior experience possess *chunks* representing functional units (Chase & Simon, 1973; Larkin et al., 1980b). Chunks are stimulus patterns or configurations that experienced users can recognize (Newell & Simon, 1972). Chase and Simon (1973) found that experts avail of chunks to identify the functional relationships and deliver better recall performance. In contrast, novices or first-time buyers are unable to provide those functional descriptions of the product in their reviews since their knowledge of the product is not organized in the form of chunks. Expert knowledge is chunked so as to include more procedural knowledge and more knowledge about the conditions of applicability (Chi et al., 1982).

Krivec et al. (2021) found that for the reconstruction of chess moves, high-skilled chess players were significantly more accurate than the low-skilled players; the speed of information recall was also markedly higher for the high-skilled group. They concluded that users with more specialized knowledge operate faster with larger procedural chunks of procedural information than those with less knowledge.

We argue that this is also the case when users write online product reviews. Consider a user who has purchased a tent. If he has purchased and used a tent before, he will most likely access chunks that help him use the functions of the new tent. The basic procedural chunks could be for steps like i) unpacking the tent supplies, ii) laying down a ground cloth, iii) inserting the tent poles through the frame, iv) raising the tent, v) hammering in the tent pegs, vi) setting up the rain-fly, and vii) testing the firmness of the tent in different weather conditions. A user with domain knowledge will possess compound procedural chunks that contain one or more basic chunks (Krivec et al., 2021). For example, a compound chunk could consist of steps iii, iv, and v. When this experienced user posts a review, he will be able to quickly recall compound chunks like this together as one unit and provide a more comprehensive and accurate evaluation of the tent's functional performance than a first-time buyer who does not have access to such compound chunks.

When examining the influence of product exposure on review helpfulness, it is therefore important to take into account the domain knowledge of the reviewer. It is this interplay of three factors – exposure time, utilitarian information, and domain knowledge – and the resulting influence on review helpfulness that we want to explore in this study. Exposure time helps users to highlight the utilitarian (functional) aspects of the product in their reviews, while prior experience with the product helps users organize their knowledge in the form of chunks, which helps them evaluate the product efficiently and accurately. While we will argue that product exposure positively influences review helpfulness, the question that arises is how much of that effect is mediated by utilitarian information. And does the influence of product exposure on review helpfulness get moderated by prior domain knowledge, given that experts already possess large chunks of knowledge that facilitate writing about the product's functional aspects? These are the major questions we address in this study. Figure 1.1 presents our research model.

Figure 1. 1 Research Model



Product Exposure and Review Helpfulness

As discussed above, with longer product exposures, customers become more familiar with the product. The increased product familiarity enhances users' understanding of their products so that they are more likely to provide helpful product descriptions. Alba and Hutchinson (1987) find that as product familiarity increases, consumers' ability to isolate the most critical and task-relevant information and accurately elaborate information is enhanced. The knowledge-assembly theory (Hayes-Roth, 1977) suggests that as product familiarity increases with learning, individuals' knowledge structures will change from a collection of independent parts to an integrated and comprehensive memory. According to Conover (1982), users' knowledge dimensionality of products will increase as users' familiarity increases, and users will tend to use more specific and concrete characteristics to describe their products. Several other studies also find that consumers with high product familiarity are more capable of extracting relevant product information while ignoring irrelevant information (Johnson & Russo, 1984; Johnson, 1984; Larkin et al., 1980a).

In summary, as product familiarity increases over the accumulated product exposure, users tend to have more relevant, accurate, and comprehensive product knowledge so that they are more likely to generate helpful product evaluations. In the online market, people express product evaluations through online reviews. Hence, it is reasonable to infer that as a user gets more exposure to a product, the helpfulness of the review she posts will also increase. We, therefore, propose the following hypothesis:

H1: Reviews by customers with longer product exposures are likely to receive more votes on review helpfulness.

Utilitarian Information

Extant studies have found that users' knowledge and perception of products are not static. Instead, they change along with the users' product experience. Several studies unearth a consistent dynamic pattern of user experience over time. They find that in the initial experience phase, customers' experience tends to focus on the hedonic aspects of a product, such as its appearance, color, and size. With prolonged experiences, users' attention gets increasingly tied to aspects of the pragmatic product attributes (Karapanos et al., 2009; Kujala et al., 2011; von Wilamowitz-Moellendorff et al., 2006), and their evaluations start placing greater emphasis on product functionality.

These findings are supported by studies on product attributes. Hassenzahl (2004) conducts one pre-use and one post-use study to reveal the impacts of user experience on hedonic and pragmatic product features. The results show that user experience has more influence on pragmatic attributes than on hedonic attributes. Hassenzahl et al. (2000) also reveal a strong correlation between post-use ratings of pragmatic quality and product appeal with appropriate product interactions. Overall, user experience has more influence on pragmatic/utilitarian product features than on hedonic product features.

The extant literature contends that users' assessments of products shift over actual product use experience, with users' attention shifting from the hedonic product attributes to the utilitarian product attributes such as functionality. Thus, we propose the following hypothesis:

H2: Reviewers with longer product exposures tend to evaluate more on the basis of utilitarian attributes.

Although hedonic attributes have proved to be an essential predictor of customer satisfaction and choice (Hassenzahl, 2001; Zauberman et al., 2006), they can be easily obtained by readers from alternative sources such as product descriptions, images, and videos posted online by sellers. On the other hand, utilitarian attributes are derived from actual interactions with products (Hirschman & Holbrook, 1982; Holbrook & Hirschman, 1982). Thus, it is difficult for potential buyers to gather this knowledge from available information sources other than online reviews. For example, customers can know a tent's appearance from its product images and size from product descriptions. However, they can only know the performance of this tent's waterproof function from the actual product experience shared by reviewers in their online reviews. Accordingly, information on utilitarian attributes should be perceived as more helpful than that on hedonic attributes due to its rareness.

Second, descriptions of utilitarian attributes tend to be objective in nature. Hedonic information is mainly derived from sensations and sentiments such as fun, joy, and excitement (Hirschman & Holbrook, 1982; Strahilevitz & Myers, 1998). Therefore, evaluations of hedonic features are based on subjective judgments, which may create bias. For example, one may feel a jacket is stylish while the other may think it looks unremarkable. In contrast, descriptions of a product's utilitarian attributes should be more objective since they are derived from the functions delivered by the product. When a reviewer evaluates the warmth of a jacket, she would usually talk about at what temperature the jacket can still keep her warm, which is an objective dimension. In other words, users tend to follow consistent criteria to evaluate the functions. As another example, a user may evaluate the waterproofing function by a tent's performance during heavy rain, where rain is an objective weather condition that everyone understands and agrees upon. On the other hand, hedonic dimensions, such as the outlook of a jacket and the color of a tent, highly depend on subjective judgments.

Because utilitarian attributes convey evaluation information regarding a product's functionality, we believe that they will be perceived by online review readers to be more helpful than hedonic attributes. Hence, we propose the following hypothesis:

H3: Reviews that contain evaluations on the basis of a product's utilitarian attributes tend to receive more helpfulness votes.

Overall, Hypotheses 1, 2, and 3, taken together, suggest that utilitarian information in a review is a potential mediator of product exposure.

The Moderating Role of Domain Knowledge

Prior literature has thoroughly investigated the effects of consumers' domain knowledge. As Alba and Hutchinson (1987) argue, one of the main effects of domain knowledge is reducing the cognitive effort to decrease performance time without losing any performance quality. Other studies endorse a similar conclusion that the amount of cognitive effort required to achieve a particular level of comprehension is lower for experts than for novices (Britton et al., 1978; Johnson & Kieras, 1983). The acceleration of comprehension for experts is because their domain knowledge – organized as chunks – enables them to select relevant information while ignoring irrelevant ones (Johnson & Russo, 1984; Johnson, 1984; Larkin et al., 1980a). These chunks help the experts articulate the product's functionality and applicability in the reviews they post online.

Because of this, experts can process and comprehend product information more efficiently and produce the helpful reviews faster than novices.

Consumers' domain knowledge is often treated as a moderator in the process of learning and evaluation in prior studies (Cheung et al., 2012; Maheswaran, 1994; Park & Kim, 2008). For example, Maheswaran et al. (1996) find that message repetition enhances novices' learning of the content and the favorability of their evaluations. However, experts' learning is mostly not influenced by repeating message information. Similarly, we propose that although product exposure enhances review helpfulness, this positive effect decreases when reviewers possess high levels of domain knowledge. The prior domain knowledge that such reviewers have is organized as compound procedural chunks (Krivec et al., 2021), which accelerate their information processing, and thus they are able to write helpful reviews with a shorter product exposure than what would be needed for reviewers without domain knowledge. Therefore, we argue that the positive effect of exposure on review helpfulness is moderated by the reviewer's domain knowledge:

H4: Reviewers' domain knowledge moderates the positive relationship between product exposure and review helpfulness.

Research Context and Data

We collected data from an e-commerce platform specializing in outdoor equipment, such as tents and snowboards, and apparel and footwear, such as snow jackets and running pants. We merged a sales dataset and a product review dataset to create the final dataset for the main analyses. The sales data contains information related to orders, including order ID, customer ID, product ID, product group, merchandise group, shipping date, shipping fee, price, and brand. It contains 352,629 transaction records from January 2014 to September 2015. The review data contains review information, including review ID, customer ID, product ID, review date, helpfulness vote, rating, review text, and self-reported familiarity. The last measure is a categorical variable containing familiarity levels (i.e., how many times the reviewer had used this product) voluntarily reported by reviewers when they post reviews. When reviewers posted reviews on our target platform, they were asked to indicate their familiarity levels with the reviewed product by selecting one of the four options: "I've put it through the wringer," "I've used it several times," "I've used it once or twice and have initial impressions," and "I returned this product before using it." We match these two datasets using product ID and customer ID with a Python program. The initial integrated dataset contains 12,010 records.

We first remove the duplicate reviews in which the reviewers use the same reviews for different purchases. The data size reduces to 11,338. We then eliminate the records that have a time difference between the review date and the shipping date that are longer than one year. We believe that it is rare that a user will post her review after receiving the product for more than one year (95^{th} percentile). The final dataset contains 10,815 records with order dates from January 2014 to September 2015. The reviews are posted by 4,095 reviewers and correspond to 4,706 products.

Product Exposure

It is challenging to capture the repetitions of a stimulus (i.e., a product) in the online business domain. In this study, we measure a customer's product exposure by the number of days between the order shipping date and the review posting date. The logic is straightforward: the longer the time difference, the more likely the customer has greater exposure to the product before posting the review.

To verify the validity of our proposed measurement, we compare product exposure with reviewers' self-reported product familiarity levels when they post reviews. The results of descriptive analysis and ANOVA test indicate a clear pattern that a higher self-reported product familiarity level corresponds to a longer average time difference between the shipping date and the review date.

The average time differences between shipping date and review date in days for these four levels are 86.65, 59.08, 37.49, and 25.13, correspondingly. The Tukey pairwise comparisons further validate that the differences of the average time difference between experience level groups are statistically significant. The details of the analyses are shown in Appendix I. The familiarity levels for the reviewed products may also be affected by users' past experience within the same product group. To eliminate the influence of past experience, we conduct a robustness check on the subsample that only contains the first-time purchases for reviewers. A reviewer will be identified as a first-time buyer if she has no past purchases within the same product group as the reviewed product. The subsample contains 8,253 reviews. The results in Appendix II are consistent with those for the full data. When the past experience's influence is removed, the familiarity levels should be exclusively formed by the product exposure time.

Utilitarian Information

This section discusses how we extract descriptions of utilitarian product attributes in a review. We adopt the Topic Modeling approach to identify the latent topics embedded in the review contents. One of the most popular topic-modeling approaches is Latent Dirichlet allocation (LDA) (Blei et al., 2003). LDA is an unsupervised method that exploits a predefined number of hidden topics from documents. We conducted the topic modeling by using the LDA implementation in the MALLET (Machine Learning for Language Toolkit) package (McCallum, 2002) via a Python package called Gensim (Rehurek & Sojka, 2011). We build 24 models with topics ranging from 2 to 50, with a step equal to two.

The results show that the model with 14 topics yields the highest coherence score. Following prior studies (Buschken & Allenby, 2016; Gong et al., 2018; Lash & Zhao 2016; Puranam et al., 2017; Shi et al., 2016), we measure topics by their contributions to documents. We assume that each document contains all topics but only differs from other documents in terms of topic proportions. In other words, each review is composed of a random mixture of several topics.

Next, we give each topic a name based on our interpretations of the informative keywords, shown in the second column of Table 1.2. We name Topic 1 as “climbing experience” because several of its keywords are relevant to climbing activities. We label Topic 2 as “overall feeling” since its keywords predominately describe subjective feelings. Topic 3, “product fit,” contains several keywords that describe the size and fit of products. Topics 4 and 5 relate to “ski experience” and “camping experience.” Topic 8 specifically describes the “heat preservation function” for water bottles. Topic 9 contains keywords, such as “light” and “heavy,” that depict “product weight.” Topic 10 sketches the “waterproof function.” Topic 12 discusses “product price.” Keywords for Topic 13 depict components of products, so we name it “product design.” Finally, Topics 6, 7, 11, and 14 contain mixed and irrelevant keywords. Thus, we name them “other.”

Table 1. 2 Topic Summaries

Topic	Keywords	Name	Group
1	climb, move, month, durable, hike, trail, hiking, adventure, mile, walk	Climbing Experience	Utilitarian Information
2	feel, solid, length, strong, sturdy, hand, cheap, size, close, quick	Overall Feeling	Hedonic Information
3	small, fit, size, big, large, medium, fits_perfectly, short, tight, comfortably	Product Fit	Hedonic Information
4	ski, ride, skiing, snow, season, tour, mountain, condition, day, winter	Ski Experience	Utilitarian Information
5	camp, camping, warm, weekend, night, sleep, rain, snow, degree, trip	Camping Experience	Utilitarian Information
6	year, time, wife, update, leak, wrong, wait, plan, decide, double	Other	Other Information
7	bad, cut, tool, include, hope, attach, offer, version, screw, provide	Other	Other Information
8	cold, ice, hot, cool, day, fill, hour, drink, water_bottle, lid	Heat Preservation Function	Utilitarian Information
9	light, weight, design, lightweight, shape, material, light_weight, heavy, feature, compact	Product Weight	Hedonic Information
10	dry, wind, weather, rain, wet, room, heavy, stay, weekend, trip	Waterproof Function	Utilitarian Information
11	gear, piece, kid, friend, gift, life, guy, absolutely, device, anchor	Other	Other Information
12	price, cheap, money, expensive, quality, excellent, heavy, design, compare, brand	Product Price	Hedonic Information
13	design, pocket, strap, front, shoulder, side, feature, color, top, gear	Product Design	Hedonic Information
14	long, extra, throw, tough, eat, fact, longer, leash, type, ready	Other	Other Information

We further classify the topics into three groups, namely “utilitarian information,” “hedonic information,” and “other information.” The classification is primarily based on the names of the topics and the definition of utilitarian and hedonic product features.

Despite the fact that different researchers have different definitions of utilitarian attributes, the common emphasis is on functionality. For example, Hassenzahl (2001) argues that utility means providing relevant functionality for performing tasks. Strahilevitz and Myers (1998) define utilitarian goods are goal-oriented and used to accomplish a functional task.

Therefore, it is clear that Topic 8, "heat preservation function," and Topic 10, "waterproof function," belong to the "utilitarian information" group. From the perspective of situatedness (Hassenzahl & Tractinsky, 2006), users' perceptions of a product's functional performance are mainly formed by interacting with the product in different situations. Therefore, topics that describe activities and experiences (e.g., "climb," "ski," "camping") are likely to be associated with descriptions of the product's functional performance. From this perspective, Topics 1, 4, and 5 should also be classified into the "utilitarian information" group.

In contrast, hedonic attributes may be derived from appearance and require minimal cognitive effort for evaluation (Hassenzahl, 2004). It comprises dimensions such as innovativeness, beauty, and stimulation (Hassenzahl et al., 2000). Thus, evaluations of hedonic attributes come predominantly from multiple sensory modalities, including tastes, sounds, scents, tactile impressions, and visual images (Hirschman & Holbrook, 1982). Summarizing from previous theories, the "hedonic information" group should contain topics related to explicit product attributes, such as color, layout, and design. These product attributes can be easily observed and stimulate users' senses to create corresponding emotions. Therefore, we classify Topic 2 (overall feeling), 3 (product fit), 9 (product weight), 12 (product price), and 13 (product design) into the "Hedonic Information" group. In summary, these five topics all describe easy-to-observe product features (Topics 3, 9, 12, and 13) or subjective feelings (Topic 2). Finally, we group the four "other" topics into the "Other Information" group.

Among the 14 topics, we identify five *utilitarian information*, five *hedonic information*, and four *other information* topics. To ensure the classification's reliability, we recruited three research assistants and asked them to independently classify the eight topics into three groups based on the keywords, names, and representative reviews. Each of them was given the definition of the three topic groups. The labeled results indicate that all the three research assistants achieved a consistent classification.

Finally, we integrate the contribution of each group by adding the contributions of the topics within that group. The focal variable, *utilitarian information*, is measured as the total contribution of the topics under the "utilitarian information" group.

Variables

Another variable of interest, *domain knowledge*, is measured as the number of past reviews within the same product group for a reviewer. The number of past reviews in the same product group does not only indicate how many times the user has experienced the product in the same category but also suggests they have enough writing experience.

We include a few important control variables in the models. First, we control for the effects of product information, including price, shipping fee, purchase date, and the fixed effects of brands following prior studies (Baek et al., 1986; Chua & Banerjee, 2015; Li & Hitt, 2010; Zhu & Zhang, 2010). To control for the effect of review existing time on review helpfulness, we derive the variable *elapsed time*, which is measured by the time difference between the review date of the target review and the data collection date in the dataset in days. Furthermore, we include review characteristics – review rating, review length, and review date (Mudambi & Schuff, 2010; Yin et al., 2016).

We also apply text mining techniques and derive the review features such as subjectivity, polarity, and readability (Ghose & Ipeirotis, 2010; Yin et al., 2014). Finally, we control for the review environment by including the total number of existing reviews before the review date. The detailed operationalization of the variables used in the main analyses and summary statistics are presented in Table 1.3.

Table 1. 3 Variable Explanation

Group	Variable Name	Description	Summary Statistics Mean (SD.)
Dependent Variable	Review Helpfulness	The number of "helpful" votes	0.483 (0.890)
Focal Variable	Product Exposure	The time difference between review date and shipping date in days.	59.805 (70.565)
	Utilitarian Information	The describtion of utilitarian product features in a review. A proportion scaled from 0 to 1.	0.36 (0.05)
	Domain Knowledge	The number of past reviews within the same product group for a reviewer.	0.255 (0.793)
Product Information	Elapsed Time	Natural logarithm of the time difference between the review date of the target review and the review date of the last review in the dataset in days plus one.	6.218 (0.699)
	Price	Natural logrithm of Price	3.727 (1.269)
	Shipping Fee	The shipping fee for the orders	1.075 (3.864)
	Month of Purchase Date	Fixed effect of month of purchase dates	Categorical Variable
	Brand	Fixed effect of brands	Categorical Variable
Review Information	Hedonic Information	The describtion of hedonic product features in a review. A proportion scaled from 0 to 1.	0.36 (0.05)
	Rating	Review rating	4.528 (0.815)
	Review Length	Natural logarithm of the number of words in a review.	4.096 (0.819)
	Month of Review Date	Fixed effects of months of review dates	Categorical Variable
Review Text Information	Subjectivity	Subjectivity score of a review ([0,1]=[objective, subjective])	0.550 (0.149)
	Polarity	Polarity score of a review ([-1,1] = [negative, positive])	0.239 (0.194)
	Fog Index	Fog index of a review	15.429 (16.408)
	PDW	Percentage of difficult words	13.640 (6.568)
Review Environment Information	Review Vol	The number of existing reviews at the time when the target reviews were posted	14.564 (30.264)

Empirical Analyses and Results

We present our data analysis results in this section. We first test the main effect of product exposure on review helpfulness (H1). Next, we analyze how review contents change over product exposure time (H2). In the next step, we further investigate how utilitarian information affects review helpfulness (H3). Then, integrating hypotheses H1, H2, and H3, we examine the possibility of a mediation relationship. Finally, we test the moderating effect of domain knowledge on the relationship between product exposure and review helpfulness.

Impact of Product Exposure on Review Helpfulness

In this section, we test the main effect of product exposure on review helpfulness. The primary independent variable is product exposure. For review *r* issued for product *j*, we model review helpfulness as:

$$\begin{aligned} ReviewHelpfulness_{rj} &= \beta_0 + \beta_1 \mathrm{ProductExposure_{rj}} + \beta_2 ElapsedTime_{rj} + \beta_3 Price_{rj} \\ &+ \beta_4 ShippingFee_{rj} + \beta_5 Rating_{rj} + \beta_6 \mathrm{ReviewVol_{rj}} + \beta_7 ReviewLength_{rj} \\ &+ \beta_8 \mathrm{Subjectivity_{rj}} + \beta_9 Polarity_{rj} + \beta_{10} FogIndex_{rj} + \beta_{11} PDW_{rj} \\ &+ \beta_{12} \mathrm{Month\_RD}_{rj} + \beta_{13} \mathrm{Month\_PD}_{rj} + \alpha_{\mathrm{j}} + \varepsilon_{rj} \quad (1) \end{aligned}$$

In equation (1), β_1 measures the main effect of product exposure of review *r* on product *j*. β_2 to β_{13} are the coefficients of the control variables. α_{j} denotes the fixed effect of brands across reviews, and ε_{rj} represents the error term.

Bias Adjustment by Accounting for the Omitted Variables

It is common to have concerns about omitted variables for econometric estimations with secondary data. In non-experimental settings, it is nearly impossible to control for all the factors. The omitted variable bias is also the most common cause of endogeneity (Wooldridge, 2002). For example, the delay of a reviewer's plan to a ski resort might affect both the product exposure time (i.e., the ski or any relevant equipment) and the helpfulness of the review (i.e., the review helpful votes). Such information is unobservable to us and may cause bias if we do not control them in the model.

To address this concern, we employ the coefficient stability approach proposed by Oster (2019). This approach argues that the robustness of estimates to omitted variable bias can be examined by observing movements in the coefficient of the focal explanatory variable (also called the treatment) and the R-squared from a baseline model that only includes the treatment to a model that consists of a complete set of control variables. If the coefficient of the treatment only changes slightly from the baseline model with a small R-squared to the full model that has a substantial increase in the R-squared, it indicates that the estimate is robust. To set the maximum R-squared (R_{max}), Oster argues that it is reasonable to set the value as 1.3 times the full model's R-squared. Our empirical analyses report the adjusted coefficient using the coefficient stability method proposed by Oster (2019).

Empirical Results

Because the dependent variable is a count variable, we use Poisson regression. The dispersion of the Poisson regression model is around 1, so over-dispersion is not an issue.

We inspect the variance inflation factors (VIFs) for each model and find that VIFs are all well below the generally established threshold of 10 (Hair et al., 1998), which indicates that multicollinearity is unlikely to confound our findings. Model 1 in Table 1.4 tests the main effect of product exposure on review helpfulness. As the results show, product exposure has a significant positive influence on review helpfulness ($\alpha = 0.00051, p < 0.05$), thus supporting H1. The coefficient stability results for product exposure are reported in Table 1.5 column 1. The results show that the adjusted coefficient is still positive; thus, it consolidates the finding that the effect of product exposure on review helpfulness is positive.

Table 1. 4 Poisson Regression Results

Variables	Model 1	Model 2	Model 3
Intercept	-8.486 (1.033) ***	-8.419 (1.058) ***	-8.403 (1.058) ***
Product Exposure	0.00051 (0.00022) *	0.00044 (0.00022) *	0.00065 (0.00023) **
Utilitarian Information		0.9423 (0.3422) **	0.95 (0.3424) **
Hedonic Information		-0.7159 (0.3318) *	-0.7101 (0.3316) *
Domain Knowledge			0.07361 (0.02016) ***
Exposure * Knowledge			-0.00074 (0.00028) **
Elapsed Time	0.4162 (0.02631) ***	0.4139 (0.02632) ***	0.4103 (0.02634) ***
Price	-0.00622 (0.01656)	-0.00936 (0.01661)	-0.00766 (0.01666)
Shipping Fee	0.00272 (0.00349)	0.00237 (0.00349)	0.0025 (0.00349)
Rating	0.1791 (0.01994) ***	0.179 (0.01999) ***	0.1777 (0.02) ***
Review Length	0.7021 (0.0228) ***	0.7022 (0.02283) ***	0.6994 (0.02285) ***
Subjectivity	-0.04645 (0.1285)	-0.0768 (0.1285)	-0.07073 (0.1285)
Polarity	-0.1602 (0.1019)	-0.149 (0.1019)	-0.1498 (0.102)
Fog Index	-0.00057 (0.00074)	-0.00057 (0.00074)	-0.00052 (0.00074)
PDW	0.0143 (0.00277) ***	0.01435 (0.00276) ***	0.01425 (0.00277) ***
Review Vol	0.00109 (0.00052) *	0.00109 (0.00053) *	0.00111 (0.00053) *
Month _RD	Yes	Yes	Yes
Month _PD	Yes	Yes	Yes
Brand	Yes	Yes	Yes
AIC	19,192.02	19,169.56	19,161.97
Log likelihood	-9,286.011	-9,272.779	-9,265.985
Dispersion	1.405632	1.400089	1.398973
N	10,815	10,815	10,815

*** p<0.001, ** p<0.01, * p<0.05

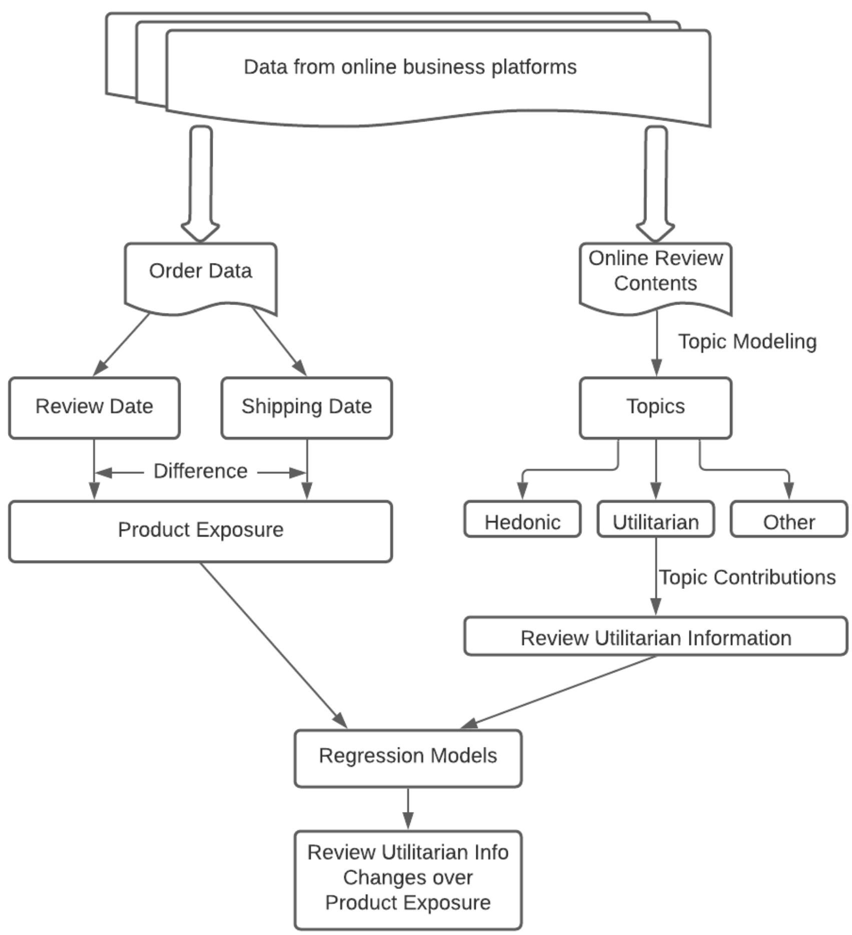
Table 1. 5 Coefficient Stability

	Product Exposure	Utilitarian	Hedonic
Adjusted Beta	0.000245	0.524	-0.413
Controlled Coefficient	0.000135	0.392	-0.302
Controlled R Square	0.173	0.174	0.174
Max R Square	0.2249	0.2262	0.226

Impact of Product Exposure on Utilitarian Information

As illustrated in the previous section, we measure utilitarian information by adopting the topic modeling method. Now, to capture the change of utilitarian information over product exposure time, we include the factors in econometric models, where the dependent variable is utilitarian information. We isolate the effects of product exposure time on utilitarian information by controlling for possible influential factors and handling omitted variables. Figure 1.2 shows how we capture the change of utilitarian information over product exposure time.

Figure 1. 2 Framework for Utilitarian Information Changes over Product Exposure



With the help of the proposed framework, we can now examine the effect of product exposure on the utilitarian information (H2). Specifically, we regress the utilitarian information on a set of explanatory variables. The dependent variable is measured as the total contribution of the topics belonging to the "utilitarian information" group. The independent variable of interest is product exposure. Additionally, we control for the same set of variables as in the previous models:

$$Review\ Utilitarian\ Information_{rj} = \gamma_0 + \gamma_1 \text{ProductExposure}_{rj} + \gamma_2 ElapsedTime_{rj} + \gamma_3 Price_{rj} + \gamma_4 ShippingFee_{rj} + \gamma_5 \text{Rating}_{rj} + \gamma_6 \text{ReviewVol}_{\text{rj}} + \gamma_7 ReviewLength_{rj} + \gamma_8 \text{Subjectivity}_{rj} + \gamma_9 Polarity_{rj} + \gamma_{10} FogIndex_{rj} + \gamma_{11} PDW_{rj} + \gamma_{12} \text{Month\_RD}_{rj} + \gamma_{13} \text{Month\_PD}_{rj} + \theta_{\text{j}} + \varphi_{rj} \qquad (2)$$

In equation (2), γ_1 measures the main effect of product exposure of review *r* on product j. γ_2 to γ_{13} are the coefficients of control variables introduced in the previous section. θ_{j} denotes the fixed effect of brands across reviews, and φ_{rj} represents the error term. We adopt the Tobit regression method because our dependent variable is measured as a proportion bounded between 0 and 1.

Empirical Results

The results of the Tobit regression model of the utilitarian information in product reviews are presented in Table 1.6. Notably, we also investigated the effect of product exposure on the hedonic information since it is the other critical product feature type and extracted from the same process as the utilitarian information. As shown in Table 1.6, product exposure positively affects utilitarian information ($\alpha = 0.00003, p < 0.001$). The positive coefficient indicates that the descriptions of utilitarian product attribute in a review increase with product exposure, which supports H2. On the contrary, the effect of product exposure on the hedonic information is negative and significant. The coefficient stability estimations indicate that omitted variables in both models do not bias the effects.

Table 1. 6 Tobit Regression Results

Variables	Model 1 (Utilitarian)	Model 2 (Hedonic)
Intercept	0.314 (0.02136) ***	0.4508 (0.02262) ***
Product Exposure	0.00003 (0.00001) ***	-0.00004 (0.00001) ***
Elapsed Time	0.00094 (0.00066)	-0.0009 (0.0007)
Price	0.00003 (0.00046)	-0.00309 (0.00048) ***
Shipping Fee	0.00026 (0.00011) *	-0.00019 (0.00011)
Rating	-0.0011 (0.00051) *	-0.00258 (0.00054) ***
Review Length	0.0001 (0.00058)	0.0013 (0.00061) *
Subjectivity	0.0117 (0.00299) ***	-0.0115 (0.00316) ***
Polarity	-0.00464 (0.00241)	0.00751 (0.00256) **
Fog Index	-0.00006 (0.00003) *	0.00003 (0.00003)
PDW	0 (0.00006)	0.00016 (0.00007) *
Review Vol	-0.00002 (0.00001)	-0.00004 (0.00002) *
Month _RD	Yes	Yes
Month _PD	Yes	Yes
Brand	Yes	Yes
R Square	0.3050288	0.29006
N	10,815	10,815

*** p<0.001, ** p<0.01, * p<0.05

Coefficient Stability	Product Exposure	Product Exposure
Adjusted Beta	0.000028	-0.000037
Controlled Coefficient	0.0000269	-0.0000353
Controlled R Square	0.305	0.29
Max R Square	0.3965	0.377

Impact of Utilitarian Information on Review Helpfulness

In this section, we examine the relationship between utilitarian information and review helpfulness, and test H3. Compared to equation (1), this model further includes the main effects of utilitarian information and hedonic information. The remaining variables are consistent with equation (1). As with the previous models, we use the Poisson regression model.

Omitted Variable Bias

We handle potential omitted variable bias for product exposure by using the coefficient stability method (Oster, 2019). We apply the same method for the utilitarian and hedonic information for addressing omitted variable issues.

Empirical Results

The results are shown in Model 2 in Table 1.4. Compared to Model 1 in Table 1.4, Model 2 includes two additional variables: utilitarian and hedonic information. The utilitarian information's significant coefficient ($\alpha = 0.94, p < 0.01$) indicates that the utilitarian information increases review helpfulness, which supports H3. On the other hand, the hedonic information negatively affects review helpfulness ($\alpha = -0.72, p < 0.05$). The coefficient stability results are reported in columns 2 and 3 of Table 1.5. The results confirm the relationship found in Table 1.4.

Mediation Test

After including the utilitarian and hedonic information, the decreased magnitude of the product exposure coefficient can provide some evidence of a mediation relationship. To unveil the significance of the mediating effect, we conduct the mediation analysis using the counterfactual framework (Pearl, 2001; Robins & Greenland, 1992). The counterfactual framework allows clear decompositions of direct and indirect effects and addresses the problems associated with the Baron and Kenny approach (1986), which has been adopted recently by several researchers. What we adopt is the version proposed by Valeri and VanderWeele (2013) because their framework supports Poisson regression models.

In the mediation relationship analysis, we can decompose the total effect (TE) into the natural direct effect (NDE) and the natural indirect effect (NIE). TE refers to how much the outcome would change overall for a change in the treatment from its reference level (a0) to treatment level (a1). NDE expresses how much the outcome would change if the treatment changes from level a0 to a1, while the mediator is held at each individual's natural value taken under treatment level a0. NIE represents the effect of allowing each individual's mediator to change from reference level a0 to treatment level a1.

In the model, we set the reference level (a0) of the treatment to the 10^{th} percentile of product exposure and a1 to its 90^{th} percentile. Because our data is not from a controlled experiment, we also include the control variables in the previous estimations. All of them are controlled at their average levels.

Table 1. 7 Mediation Analysis Results: Utilitarian

Effect	Est. [95% CI]	Sig.
Total Effect (TE)	0.110 [0.042, 0.178]	**
Natural Direct Effect (NDE)	0.104 [0.036, 0.172]	**
Natural Indirect Effect (NIE)	0.006 [0.002, 0.009]	**
*All effects are conditional on covariates at their average levels.		
a0 = 10th percentile; a1 = 90th percentile		
Abbreviations: Est.: estimate, CI: confidence interval; Sig.: significant level		
*** p<0.001, ** p<0.01, * p<0.05		

The findings of the mediation analysis with the utilitarian information as the mediator are summarized in Table 1.7. As the results suggest, the significant natural indirect effect (NIE) proves a significant mediation relationship. We also test the model with the hedonic information as the mediator. The results indicate a significant mediation effect.

The Moderating Effect of Domain Knowledge

Finally, we test the moderating effect of domain knowledge for the relationship between product exposure and review helpfulness. Specifically, we further include domain knowledge and its interaction term with product exposure in the model. Model 3 of Table 1.4 contains the results. The significant negative coefficient of the interaction term between product exposure and domain knowledge ($\alpha = -0.00074, p < 0.01$) indicates that reviewers with a higher level of domain knowledge do not need to have a longer product exposure to produce a review, thus supporting H4. More importantly, the coefficient of product exposure in Model 3 ($\alpha = 0.00065, p < 0.01$) is still significantly positive, and its magnitude is larger than that of the corresponding coefficient in Model 2. Such a result implies that product exposure is especially important when the reviewer has no domain knowledge, which further confirms the main effect of product exposure on review helpfulness.

First-time Buyers versus Users with Domain Knowledge

To better understand the influence of the interplay of exposure time, utilitarian information, and domain knowledge on review helpfulness, we conduct a post-hoc analysis by comparing the results between two subsamples: first-time buyers and experienced buyers. As we introduced in the earlier section, the first-time buyers' subsample contains all the reviews that are written by the reviewers who have no purchase experience for the same product group as the reviewed products. On the other hand, the experienced buyers' subsample has all the reviews written by the reviewers with at least one past purchase experience.

This section reports the Poisson regression models' results for these two subsamples and discusses the differences. The results are summarized in Table 1.8. As can be seen, the effect of product exposure is significant for first-time buyers, but not significant for experienced buyers. It indicates that the positive effect of product exposure on review helpfulness is only significant when reviewers have no domain knowledge. However, when the reviewers have gathered domain knowledge through prior experience, they will not need an extended exposure time to have a helpful review. It is because domain knowledge can help reviewers organize the product knowledge in chunks (Chase & Simon, 1973: Larkin et al., 1980b).

Table 1. 8 First-time buyers versus experienced buyers

	First-time Buyers		Experienced Buyers	
Variables	First-time Buyers 1	First-time Buyers 2	Experienced Buyers 1	Experienced Buyers 2
(Intercept)	-8.174 (1.042) ***	-8.023 (1.076) ***	-23.94 (1518)	-24.19 (1524)
Product Exposure	0.00063 (0.00024) *	0.00056 (0.00024) *	0.00064 (0.00055)	0.00055 (0.00055)
Utilitarian Information		0.9529 (0.3974) *		0.889 (0.756)
Hedonic Information		-0.9373 (0.3814) *		-0.2282 (0.75)
Control Variables	Yes	Yes	Yes	Yes
N	8,253	8,253	2,562	2,562

*** p<0.001, ** p<0.01, * p<0.05

Robustness Checks

In this section, we conduct robustness checks for our main estimation results. In the interest of space, the result tables for this section only present the effects of the focal variables.

First, we restrict to a one-year product exposure time window in our main models. We conduct robustness checks with three different window lengths (full data, three quarters, and two quarters) to verify that our results are robust across different time windows. We decide to run the models by using three different time windows to validate the estimation results in longer and shorter time windows. The full data contains all the matched reviews from the initial datasets. It spans over 1,148 days. The results in Table 1.9 indicate that the effects of product exposure and utilitarian information remain consistent.

The moderating effect of the domain knowledge is negative and significant for the full data, but not significant for the three-quarters and two-quarters scenarios. We believe that domain knowledge can accelerate users' comprehension process of their products because it helps users isolate relevant and important information while ignoring noises (Johnson & Russo, 1984; Johnson, 1984; Larkin et al., 1980a) in a shorter time. However, it still requires a certain exposure time for users to experience the products before they can have helpful reviews, even for those with domain knowledge. Therefore, the moderating effect will not be significant when we restrict the product exposure time to a shorter length.

Table 1. 9 Robustness Check 1: Time Windows

	Full		**Three Quarters**		**Two Quarters**	
Variables	Full 1	Full 2	Three Quarters 1	Three Quarters 2	Two Quarters 1	Two Quarters 2
(Intercept)	-8.544 (1.031) ***	-8.461 (1.056) ***	-8.48 (1.034) ***	-8.459 (1.059) ***	-8.59 (1.037) ***	-8.645 (1.064) ***
Product Exposure	0.0003 (0.00014) *	0.00026 (0.00014) †	0.00123 (0.00029) ***	0.00117 (0.00029) ***	0.0008 (0.00035) *	0.00073 (0.00041) †
Domain Knowledge	0.0634 (0.0190) ***	0.06385 (0.0190) ***	0.06794 (0.02136) **	0.0678 (0.02131) **	0.05555 (0.02298) *	0.05354 (0.02296) *
Exposure × Knowledge	-0.00044 (0.00017) *	-0.00045 (0.00017) **	-0.00046 (0.00035)	-0.00045 (0.00035)	0.00002 (0.00045)	0.0001 (0.00045)
Utilitarian Information		0.9216 (0.3344) **		0.9498 (0.3462) **		1.03 (0.3551) **
Hedonic Information		-0.7244 (0.3267) *		-0.6328 (0.3343)		-0.5202 (0.3402)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
N	11,338	11,338	10,480	10,480	9,954	9,954

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1

Second, product exposure time is measured as the time difference between shipping and review dates. We do not have specific usage starting dates because it is an empirical study with secondary online data. Therefore, it is reasonable to question how to control the possible time gap between the shipping date and the first usage date. Our main estimations adapt the coefficient stability approach to account for possible omitted variables. In this section, we further check the robustness from various perspectives. According to the characteristics of products, some of them can be assessed immediately after users receive them, such as foods, beverages, and accessories. On the other hand, other products may need to be evaluated at some events or during a certain time, such as snowboards, tents, and ski shoes. As a result, users are likely to use those products later after the deliveries. We conduct a subsample analysis only on those products that can be evaluated immediately. The subsample contains 6,846 reviews. By doing so, we can lower the possibility of delayed use. The results in Table 1.10 show consistent results.

In addition to identifying possible delays in usage from product characteristics, we further search based on the purchase dates. Some products are primarily used during certain periods. For example, winter gears such as snowboards can only be used in the snow season (i.e., November to March). If the purchase dates of the winter gears are not in this period, they will likely be used later after deliveries. We first identify offseason purchases based on products. We consider two types of products: winter and summer products, such as sandals and bikinis. In total, we identify 193 offseason purchases. We remove them and conduct a subsample analysis. The results are consistent with the main models. Furthermore, some products may not have strong seasonality. However, they can be used during a specific time. We further identify offseason purchases from review contents. Some reviews explicitly mention the use time. We extract keywords that reveal use time, such as seasons and months, using the text mining technique. Then, we compare those keywords with the purchase dates to identify offseason purchases. We find 95 offseason purchases with this approach. Similarly, we remove them from the data and conduct a subsample analysis. The results are also consistent. The consistent results from the three subsample analyses show that the effects in the main estimations are not biased by possible delays in product usage.

Table 1. 10 Robustness Check 2: Product Exposure Measure Validity

	Immediate Evaluation		No Offseason		No Review Offseason	
Variables	Immediate Evaluation 1	Immediate Evaluation 2	No Offseason	No Offseason	No Review Offseason 1	No Review Offseason 2
(Intercept)	-6.531 (0.7779) ***	-7.075 (0.8313) ***	-8.405 (1.033) ***	-8.367 (1.058) ***	-8.495 (1.033) ***	-8.403 (1.059) ***
Product Exposure	0.00059 (0.0003) *	0.00053 (0.0003) †	0.00074 (0.00023) **	0.00067 (0.00023) **	0.00071 (0.00023) **	0.00064 (0.00023) **
Domain Knowledge	0.06372 (0.02363) **	0.06285 (0.02366) **	0.07363 (0.0203) ***	0.07378 (0.0202) ***	0.07132 (0.0203) ***	0.07161 (0.0203) ***
Exposure * Knowledge	-0.00072 (0.00036) *	-0.00075 (0.00036) *	-0.00073 (0.00028) **	-0.00073 (0.00028) **	-0.00074 (0.00028) **	-0.00074 (0.00028) **
Utilitarian Information		1.412 (0.4783) **		0.9617 (0.346) **		0.9476 (0.3454) **
Hedonic Information		0.1039 (0.46)		-0.6668 (0.3345) *		-0.7821 (0.3341) *
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
N	6,846	6,846	10,622	10,622	10,720	10,720

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1

Third, in our main models, we assume each review is written at one single point. However, some reviews may contain updates. The target platform does not have the update reviews function. But it allows reviewers to make revisions to their original reviews. To remove the effects of the updated reviews, we identify them using the text mining technique. Specifically, we extract relevant keywords, such as "update," "revision," and "edition," from review contents. We found 48 reviews that explicitly mention that there is updated information in the reviews. We remove these reviews and conduct a subsample analysis. The results in Table 1.11 are consistent.

Table 1. 11 Robustness Check 3: No updated reviews

Variables	No Updated Reviews 1	No Updated Reviews 2
(Intercept)	-8.469 (1.033) ***	-8.426 (1.058) ***
Product Exposure	0.00072 (0.00023) **	0.00065 (0.00023) **
Domain Knowledge	0.06907 (0.02067) ***	0.06961 (0.02062) ***
Exposure * Knowledge	-0.00072 (0.00028) *	-0.00072 (0.00028) *
Utilitarian Information		0.9773 (0.345) **
Hedonic Information		-0.6853 (0.3344) *
Control Variables	Yes	Yes
N	10,767	10,767

*** p<0.001, ** p<0.01, * p<0.05

Finally, we account for the product types in the main models by controlling the brands' fixed effects. However, different product types can be from the same brand. To further control the product type effects, we conduct a robustness check by controlling for fixed effects of the product merchandise groups. The results in Table 1.12 are consistent with the main estimations.

Table 1. 12 Robustness Check 4: Product merchandise group

Variables	Merchandise Groups 1	Merchandise Groups 2
(Intercept)	-7.898 (0.6317) ***	-8.155 (0.6621) ***
Product Exposure	0.00062 (0.00023) **	0.00055 (0.00023) *
Domain Knowledge	0.04376 (0.01839) *	0.04274 (0.01829) *
Exposure * Knowledge	-0.00053 (0.00027) *	-0.00051 (0.00027) †
Utilitarian Information		1.178 (0.3828) **
Hedonic Information		-0.3168 (0.3636)
Control Variables	Yes	Yes
Merchandise Groups	Yes	Yes
N	10,815	10,815

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1

Discussion and Managerial Implications

This essay investigates the interplay of product exposure time, utilitarian information, and domain knowledge, and its effect of review helpfulness. In general, we find that product exposure positively impacts review helpfulness, and utilitarian information mediates this effect. Moreover, domain knowledge negatively moderates the positive effect of product exposure on review helpfulness.

We apply a Poisson regression model to investigate the effects of product exposure on review helpfulness. The positive effect of product exposure on review helpfulness indicates that more prolonged product exposure leads to more helpful online reviews. Longer product exposure increases users' familiarity with the products, which leads to a better understanding. When users have a more comprehensive understanding of the products they purchase, their evaluations will be more helpful to future customers.

To study the role of utilitarian information, we apply machine learning to extract review topics and then analyze how they evolve across product exposure by econometric modeling. The results from the Tobit regression analysis show a significant positive relationship between product exposure and utilitarian information. It means that with longer product exposure times, descriptions of utilitarian product attributes in a review increase.

We find a significant positive effect of utilitarian information on review helpfulness, indicating that utilitarian product information is perceived as more helpful. The results together reveal a significant mediation relationship. Specifically, the positive effect of product exposure on review helpfulness is positively mediated by utilitarian information. The counterfactual framework verifies a significant mediating effect. Notably, we also investigate the role of hedonic information in our research framework. Hedonic information decreases as product exposure increases and negatively affects review helpfulness. We believe the negative effects of hedonic information are due to its shallow and explicit characteristics.

Finally, we find that the positive effect of product exposure on review helpfulness is mitigated by consumers' domain knowledge. More specifically, if a user has prior knowledge of a particular domain, the effect of product exposure on review helpfulness will decrease. In other words, domain knowledge can enable users to comprehend and evaluate products within a shorter period of product exposure. The post-hoc analysis on first-time and experienced buyers further proves that product exposure's effect on review helpfulness is only significant when reviewers have no domain knowledge. It is because experienced users can organize the product knowledge in large-sized procedural chunks with the help of domain knowledge. Therefore, they do not need extended exposure time to form utilitarian information (Chase & Simon, 1973; Krivec et al., 2021; Larkin et al., 1980b).

Although prior studies have identified various influential factors for review helpfulness, temporal factors have drawn little attention, especially product exposure. Moreover, to the best of our knowledge, this study is the first attempt to examine the dynamic pattern of product evaluations over time in an online setting. Previous user experience studies have discovered a consistent changing pattern of product evaluations over time in offline environments. Our findings on the effect of product exposure on utilitarian information and hedonic information confirm a similar pattern for online reviews. Furthermore, this study expands the research territory for user experience research to online reviews. From the perspective of utilitarian versus hedonic product features, previous studies have largely investigated how they affect factors such as product evaluation. However, no study that has studied how different product features' information changes over time. We further extend the relationship by investigating how this changing pattern positively affects review helpfulness. Finally, we introduce domain knowledge and study its interplay with product exposure and utilitarian information. To the best of our knowledge, this study is the first attempt to investigate such an interplay and its impact on review helpfulness. The results can broaden our understanding of product evaluation and review helpfulness.

The findings of our study have practical business implications. First, we find that review helpfulness increases as reviewers are exposed to the products for a more extended period. Many companies send review requests to their customers too soon, such as three days after customers receive their products. Based on our findings, this is not a wise strategy. Instead, we recommend that companies delay their review requests so that customers can have more time to experience their products and, consequently, post helpful reviews. For sellers who aim to procure helpful reviews from the customers as early as possible, we suggest that they target customers with a high level of domain knowledge, i.e., those who have posted reviews for products in the same product category before. Reviews by these customers after a shorter product exposure may still contain helpful information. The post-hoc analysis also supports this conclusion.

We also find that the positive effect of product exposure on review helpfulness arises because of the increase of utilitarian information in a review. Therefore, an online seller or platform might utilize our findings in developing writing guidelines to encourage users to share more on their evaluations of a product's utilitarian attributes. For example, they can post statements before the review section to remind users about their product experience with respect to product utilitarian performance, such as "Could you recall which functions of your product impressed you the most?" Sellers can also send one or two emails to ask about their experience with some of the competitive functions of their products before sending review requests, such as "We hope you are enjoying a cozy camping with our patented waterproof technique." Such emails can remind customers to experience key functions of products, thereby producing helpful reviews. The same results can also help develop companies' marketing strategies. Based on our findings, we recommend the companies emphasize their products' functional performance in their marketing channels since audiences prefer to learn utilitarian product information over hedonic or other information.

Finally, our findings provide online businesses with a convenient and reliable indicator of review helpfulness: product exposure. Platforms always want to arrange helpful reviews in prioritized positions to provide more information to their audiences. Some platforms label useful reviews as "featured reviews." However, since the accumulation of helpful votes needs time, it is important to obtain early review helpfulness predictors. The results confirm that product exposure, which can be easily gathered by online platforms right after a review is posted, is a valid and strong indicator of review helpfulness.

Conclusions, Limitations, and Future Directions

Understanding the role of users' product exposure in product evaluation generation and decision-making has always been considered a crucial research topic in various areas, such as human-computer interaction and marketing. Even though e-commerce platforms and online reviews dominate the business world, product exposure for online customers has drawn limited attention. In this essay, we investigate the interplay of product exposure time, utilitarian information, and domain knowledge and its impact on review helpfulness. Our results suggest a mediation relationship, which is that, as users' product exposure accumulates, the descriptions of utilitarian attributes in a review increase, and, in turn, increase review helpfulness. Furthermore, domain knowledge mitigates the positive effect of product exposure on review helpfulness.

Our study suffers from a few limitations. First, the counterfactual framework reveals a significant mediation relationship. However, the regression models indicate that the mediator cannot fully mediate the effect (Baron & Kenny, 1986). Therefore, there may exist other potential mediators for the exposure-helpfulness relationship that have not yet been discovered.

Following this direction, future studies can explore other potential mediators for the exposure-helpfulness relationship.

Second, although we control for the fixed effect of brands in our models, we do not differentiate between product types, such as experience and search goods, in our model. Our target platform specializes in outdoor equipment and apparel, which creates difficulties in obtaining a clear product type classification. For example, a pair of shoes is commonly identified as a search good because it can be evaluated prior to purchase and consumption (Nelson, 1970). However, when it comes to a pair of climbing shoes, it could be an experience good since it requires a certain level of product experience. Nevertheless, we believe that product types could moderate the effect of product exposure on review content and review helpfulness. Therefore, other than controlling the brands' fixed effects in the main estimations, we also control the product merchandise groups' fixed effects in robustness checks. Despite the effort, we believe that the best approach is to test such a moderating effect directly. Future researchers can take a further step to investigate such a moderating effect of product types.

Third, we believe product exposure should affect not only the review helpfulness but also other review features such as the rating score and the sentiments in the reviews. Besides, according to the existing studies, product exposure may also influence users' overall evaluations (Fredrickson, 2000), predict future behavior (Oishi & Sullivan, 2006), and enhance customer loyalty and willingness to recommend the product to others (Kujala et al., 2011). However, these aspects have not been studied in the online business domain yet. Future studies can investigate the influence of product exposure on other variables in the context of online business.

Fourth, we only focus on the utilitarian and hedonic information in this essay. Studies in the marketing literature have investigated user experiences from several other perspectives, such as functional, form, structure, emotion, and value (Brown, 2003; Desmet & Hekkert, 2007; Hassenzahl, 2004; McGinnis & Ullman, 1992; Norman, 2004). All those perspectives have not yet been fully explored in the online business context. Even with these limitations, we believe this study provides new insights into a unique business scenario involving product exposure. We hope it encourages continued investigations of this relevant and critical issue in the online market.

Finally, our target platform does not have a function for reviewers to update their reviews. Therefore, we do not have enough reviews that contain the updated information in our data. We eliminate those updated reviews identified through review content for the robustness check, but we cannot do any analyses on them. If future studies can have enough updated reviews and the corresponding timestamps, we believe it will be interesting if they can investigate the effects of product exposure from the perspective of updated review.

References

1. Agnihotri, A., & Bhattacharya, S. (2016). Online review helpfulness: Role of qualitative factors. *Psychology & Marketing*, *33*(11), 1006-1017.
2. Alba, J. W., & Hutchinson, J. W. (1987). Dimensions of consumer expertise. *Journal of Consumer Research*, *13*(4), 411-454.
3. Anderson, R. D., Engledow, J. L., & Becker, H. (1979). Evaluating the relationships among attitude toward business, product satisfaction, experience, and search effort. *Journal of Marketing Research*, *16*(3), 394-400.
4. Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of online consumer reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce*, *17*(2), 99-126.
5. Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173.
6. Bettman, J. R., & Park, C. W. (1980). Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *Journal of Consumer Research*, *7*(3), 234-248.
7. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993-1022.
8. Botti, S., & McGill, A. L. (2011). The locus of choice: Personal causality and satisfaction with hedonic and utilitarian decisions. *Journal of Consumer Research*, *37*(6), 1065-1078.
9. Britton, B. K., Westbrook, R. D., & Holdredge, T. S. (1978). Reading and cognitive capacity usage: Effects of text difficulty. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 582.
10. Brown, D. C. (2003, January). Functional, behavioral and structural features. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Vol. 37017, pp. 895-900).
11. Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science*, *35*(6), 953-975.
12. Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems*, *50*(2), 511-521.
13. Simon, H. A., & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, *5*(1), 29-46.
14. Cheung, C. M., Xiao, B., & Liu, I. L. (2012, January). The impact of observational learning and electronic word of mouth on consumer purchase decisions: The moderating role of consumer expertise and consumer involvement. In *2012 45th Hawaii International Conference on System Sciences* (pp. 3228-3237). IEEE.
15. Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345-354.
16. Chi, M., Glaser, R., Rees, E., & Steinberg, R. J. (1982). Advances in the psychology of human intelligence.
17. Chua, A. Y., & Banerjee, S. (2015). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. *Journal of the Association for Information Science and Technology*, *66*(2), 354-362.

18. Clemons, E. K., Gao, G. G., & Hitt, L. M. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, *23*(2), 149-171.
19. Conover, J. N. (1982). Familiarity and the structure of product knowledge. *ACR North American Advances*.
20. Cordell, V. V. (1997). Consumer knowledge measures as predictors in product evaluation. *Psychology & Marketing*, *14*(3), 241-260.
21. Desmet, P., & Hekkert, P. (2007). Framework of product experience. *International Journal of Design*, *1*(1).
22. Dhar, R., & Wertenbroch, K. (2000). Consumer choice between hedonic and utilitarian goods. *Journal of Marketing Research*, *37*(1), 60-71.
23. Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*(3), 291-313.
24. Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition & Emotion*, *14*(4), 577-606.
25. Ghose, A., & Ipeirotis, P. G. (2010). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498-1512.
26. Gong, J., Abhishek, V., & Li, B. (2018). Examining the impact of keyword ambiguity on search advertising performance: A topic model approach. *MIS Quarterly (Forthcoming)*.
27. Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). Multivariate data analysis prentice hall. *Upper Saddle River, NJ*, *730*.
28. Ham, J., Lee, K., Kim, T., & Koo, C. (2019). Subjective perception patterns of online reviews: A comparison of utilitarian and hedonic values. *Information Processing & Management*, *56*(4), 1439-1456.
29. Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, *13*(4), 481-499.
30. Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human–Computer Interaction*, *19*(4), 319-349.
31. Hassenzahl, M. (2008, September). User experience (UX) towards an experiential perspective on product quality. In *Proceedings of the 20th Conference on l'Interaction Homme-Machine* (pp. 11-15).
32. Hassenzahl, M., Platz, A., Burmester, M., & Lehner, K. (2000, April). Hedonic and ergonomic quality aspects determine a software's appeal. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems* (pp. 201-208).
33. Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & Information Technology*, *25*(2), 91-97.
34. Hayes-Roth, B. (1977). Evolution of cognitive structures and processes. *Psychological Review*, *84*(3), 260.
35. Hazari, S., Bergiel, B. J., & Sethna, B. N. (2017). Hedonic and utilitarian use of user-generated content on online shopping websites. *Journal of Marketing Communications*, *23*(6), 572-591.
36. Heimbach, A. E., Johansson, J. K., & MacLachlan, D. L. (1989). Product familiarity, information processing, and country-of-origin cues. *ACR North American Advances*.

37. Heng, Y., Gao, Z., Jiang, Y., & Chen, X. (2018). Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach. *Journal of Retailing and Consumer Services*, *42*, 161-168.
38. Hirschman, E. C., & Holbrook, M. B. (1982). Hedonic consumption: emerging concepts, methods and propositions. *Journal of Marketing*, *46*(3), 92-101.
39. Holbrook, M. B., & Hirschman, E. C. (1982). The experiential aspects of consumption: Consumer fantasies, feelings, and fun. *Journal of Consumer Research*, *9*(2), 132-140.
40. Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, *48*, 17-27.
41. Huang, L., Tan, C. H., Ke, W., & Wei, K. K. (2018). Helpfulness of online review content: The moderating effects of temporal and social cues. *Journal of the Association for Information Systems*, *19*(6), 3.
42. Jacoby, J., Chestnut, R. W., & Fisher, W. A. (1978). A behavioral process approach to information acquisition in nondurable purchasing. *Journal of Marketing Research*, *15*(4), 532-544.
43. Johnson, E. J., & Russo, J. E. (1984). Product familiarity and learning new information. *Journal of Consumer Research*, *11*(1), 542-550.
44. Johnson, M. D. (1984). Consumer choice strategies for comparing noncomparable alternatives. *Journal of Consumer Research*, *11*(3), 741-753.
45. Johnson, W., & Kieras, D. (1983). Representation-saving effects of prior knowledge in memory for simple technical prose. *Memory & Cognition*, *11*(5), 456-466.
46. Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J. B. (2010). Measuring the dynamics of remembered experience over time. *Interacting with Computers*, *22*(5), 328-335.
47. Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J. B. (2009, April). User experience over time: an initial framework. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 729-738).
48. Ketelaar, P. E., Willemsen, L. M., Sleven, L., & Kerkhof, P. (2015). The good, the bad, and the expert: how consumer expertise affects review valence effects on purchase intentions in online product reviews. *Journal of Computer-Mediated Communication*, *20*(6), 649-666.
49. Kim, E. E. K., Mattila, A. S., & Baloglu, S. (2011). Effects of gender and expertise on consumers' motivation to read online hotel reviews. *Cornell Hospitality Quarterly*, *52*(4), 399-406.
50. Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, *11*(3), 205-217.
51. Krivec, J., Bratko, I., & Guid, M. (2021). Identification and conceptualization of procedural chunks in chess. *Cognitive Systems Research*, *69*, 22-40.
52. Kronrod, A., & Danziger, S. (2013). "Wii will rock you!" The use and effect of figurative language in consumer reviews of hedonic and utilitarian consumption. *Journal of Consumer Research*, *40*(4), 726-739.
53. Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. *Interacting with Computers*, *23*(5), 473-483.
54. Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of competence in solving physics problems. *Cognitive Science*, *4*(4), 317-345.

55. Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*(4450), 1335-1342.
56. Lash, M. T., & Zhao, K. (2016). Early predictions of movie success: The who, what, and when of profitability. *Journal of Management Information Systems*, *33*(3), 874-903.
57. Law, E. L. C., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009, April). Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 719-728).
58. Li, X., & Hitt, L. M. (2010). Price effects in online product reviews: An analytical model and empirical analysis. *MIS Quarterly*, 809-831.
59. Lu, S., Wu, J., & Tseng, S. L. A. (2018). How online reviews become helpful: A dynamic perspective. *Journal of Interactive Marketing*, *44*, 17-28.
60. Luo, X., & Zhang, J. (2013). How do consumer buzz and traffic in social media marketing predict the value of the firm?. *Journal of Management Information Systems*, *30*(2), 213-238.
61. Ma, X., Khansa, L., Deng, Y., & Kim, S. S. (2013). Impact of prior reviews on the subsequent review process in reputation systems. *Journal of Management Information Systems*, *30*(3), 279-310.
62. Maheswaran, D. (1994). Country of origin as a stereotype: Effects of consumer expertise and attribute strength on product evaluations. *Journal of Consumer Research*, *21*(2), 354-365.
63. Maheswaran, D., Sternthal, B., & Gürhan, Z. (1996). Acquisition and impact of consumer expertise. *Journal of Consumer Psychology*, *5*(2), 115-133.
64. McCallum, A. K. (2002). Mallet: A machine learning for language toolkit. *http://mallet.cs.umass.edu*.
65. McGinnis, B. D., & Ullman, D. G. (1992). The Evolution of Commitments in the Design of a Component. *Journal of Mechanical Design, 114*(1), 1-7.
66. Moore, S. G. (2015). Attitude predictability and helpfulness in online reviews: The role of explained actions and reactions. *Journal of Consumer Research*, *42*(1), 30-44.
67. Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS Quarterly*, *34*(1), 185-200.
68. Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, *78*(2), 311-329.
69. Newell, A., & Simon, H. A. (1972). *Human Problem Solving* (Vol. 104, No. 9). Englewood Cliffs, NJ: Prentice-hall.
70. Ngo-Ye, T. L., & Sinha, A. P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, *61*, 47-58.
71. Norman, D. A. (2004). Introduction to this special section on beauty, goodness, and usability. *Human–Computer Interaction*, *19*(4), 311-318.
72. Oishi, S., & Sullivan, H. W. (2006). The predictive value of daily vs. retrospective well-being judgments in relationship stability. *Journal of Experimental Social Psychology*, *42*(4), 460-470.
73. Olsson, T., & Salo, M. (2012, May). Narratives of satisfying and unsatisfying experiences of current mobile augmented reality applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2779-2788).
74. Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187-204.
75. Park, C. W., & Lessig, V. P. (1981). Familiarity and its impact on consumer decision biases and heuristics. *Journal of Consumer Research*, *8*(2), 223-230.

76. Park, D. H., & Kim, S. (2008). The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. *Electronic Commerce Research and Applications*, *7*(4), 399-410.
77. Pearl, J. (2001). Direct and indirect effects. In J. Breese & D. Koller (Eds.), *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence.* San Francisco, CA: Morgan Kaufmann.
78. Puranam, D., Narayan, V., & Kadiyali, V. (2017). The effect of calorie posting regulation on consumer opinion: A flexible latent Dirichlet allocation model with informative priors. *Marketing Science*, *36*(5), 726-746.
79. Raju, P. S., & Reilly, M. D. (1980). Product familiarity and information processing strategies: An exploratory investigation. *Journal of Business Research*, *8*(2), 187-212.
80. Raju, P. S. (1977). Product familiarity, brand name, and price influences on product evaluation. *ACR North American Advances*.
81. Rehurek, R., & Sojka, P. (2011). Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, *3*(2), 2.
82. Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 143-155.
83. Shi, Z., Lee, G. M., & Whinston, A. B. (2016). Toward a better measure of business proximity. *MIS Quarterly*, *40*(4), 1035-1056.
84. Staiano, J., Menéndez, M., Battocchi, A., De Angeli, A., & Sebe, N. (2012, June). UX_Mate: from facial expressions to UX evaluation. In *Proceedings of the Designing Interactive Systems Conference* (pp. 741-750).
85. Stang, D. J. (1973). Six theories of repeated exposure and affect. *JSAS Catalog of Selected Documents in Psychology*, *3*, 126.
86. Stang, D. J. (1975). Effects of" mere exposure" on learning and affect. *Journal of Personality and Social Psychology*, *31*(1), 7.
87. Strahilevitz, M., & Myers, J. G. (1998). Donations to charity as purchase incentives: How well they work may depend on what you are trying to sell. *Journal of Consumer Research*, *24*(4), 434-446.
88. Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, *18*(2), 137.
89. Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010, October). User experience evaluation methods: current state and development needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (pp. 521-530).
90. von Wilamowitz-Moellendorff, M., Hassenzahl, M., & Platz, A. (2006, October). Dynamics of user experience: How the perceived quality of mobile phones changes over time. In *User Experience-Towards a unified view, Workshop at the 4th Nordic Conference on Human-Computer Interaction* (pp. 74-78).
91. Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems*, *79*, 12-23.

92. Woo, J. B., & Lim, Y. K. (2015, September). User experience in do-it-yourself-style smart homes. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 779-790).
93. Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. MIT press.
94. Yin, D., Bond, S. D., & Zhang, H. (2017). Keep your cool or let it out: Nonlinear effects of expressed arousal on perceptions of consumer reviews. *Journal of Marketing Research*, *54*(3), 447-463.
95. Yin, D., Bond, S., & Zhang, H. (2020). Anger in consumer reviews: Unhelpful but persuasive?. *MIS Quarterly, Forthcoming, Georgia Tech Scheller College of Business Research Paper*, (3588859).
96. Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, *38*(2), 539-560.
97. Yin, D., Mitra, S., & Zhang, H. (2016). Research note—When do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth. *Information Systems Research*, *27*(1), 131-144.
98. Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2p2), 1.
99. Zajonc, R. B., Markus, H., & Wilson, W. R. (1974). Exposure effects and associative learning. *Journal of Experimental Social Psychology*, *10*(3), 248-263.
100. Zauberman, G., Diehl, K., & Ariely, D. (2006). Hedonic versus informational evaluations: Task dependent preferences for sequences of outcomes. *Journal of Behavioral Decision Making*, *19*(3), 191-211.
101. Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, *74*(2), 133-148.
102. Zou, P., Yu, B., & Hao, Y. (2011). Does the valence of online consumer reviews matter for consumer decision making? The moderating role of consumer expertise. *Journal of Computers*, *6*(3), 484-488.

Appendix A. Product exposure vs. familiarity levels

Familiarity	Average of Product Exposure	Count	Std. Dev.
I've put it through the wringer	86.65	3,093	86.43
I've used it several times	59.08	3,311	67.15
I've used it once or twice	37.49	1,539	46.88
I returned this product before using it	25.13	216	45.36

ANOVA test

	DF	**F Value**	**Pr.(>F)**
Familiarity	3	199.6	<2e-16 ***
Residuals	8155		

Tukey multiple pairwise-comparisons

	Difference	**P value**	**Sig.**
Wringer - Return Before Using	61.51689	0	***
Once/Twice - Return Before Using	12.35307	0.082	†
Several Times - Return Before Using	33.94729	0	***
Once/Twice - Wringer	-49.16382	0	***
Several Times - Wringer	-27.5696	0	***
Several Times - Once/Twice	21.59422	0	***

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1

Appendix B. Product exposure vs. familiarity levels for the first-time buyers

Familiarity	Average of Product Exposure	Count	Std. Dev.
I've put it through the wringer	89.117	2,212	87.99
I've used it several times	60.839	2,578	67.99
I've used it once or twice	38.097	1,257	47.68
I returned this product before using it	22.988	169	38.44

	DF	**F Value**	**Pr.(>F)**
Familiarity	5	139.5	<2e-16 ***
Residuals	8247		

	Difference	**P value**	**Sig.**
Wringer - Return Before Using	66.1289	0.0000	***
Once/Twice - Return Before Using	15.1089	0.0731	†
Several Times - Return Before Using	37.8509	0.0000	***
Once/Twice - Wringer	-51.0200	0.0000	***
Several Times - Wringer	-28.2781	0.0000	***
Several Times - Once/Twice	22.7420	0.0000	***

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1

Essay II:

When Perceptions Match Reality: The Influence of Hospital Medical Performance and Review Content on Online Ratings

Abstract

Online healthcare reviews have a significant impact on patients' provider selections, providers' demands, and online reputation. However, there is a lack of a comprehensive understanding of the factors that influence online healthcare reviews. Previous research has generated mixed and inconsistent results. Unlike the majority of prior studies that examine the influence of online reviews on healthcare providers' medical performance or the correlations between them, we investigate how medical performance factors affect hospitals' online ratings. We find that readmission, mortality, safety of care, and time in emergency department significantly influence a hospital's online reputation. We also extract three influential review content factors: reviewer medical knowledge, medical quality evaluations, and service quality evaluations. This is the first study to investigate the effect of a set of representative hospital medical performance factors on online ratings. Furthermore, it is the first attempt at examining the roles of reviewer medical knowledge and different types of experiential quality evaluations in the online healthcare review domain. We also find a significant influence of CMS overall quality star rating on a hospital's online reputation. The findings provide valuable inputs into a hospital's marketing strategies and have important managerial implications for providers, patients, and online platforms.

Keywords: online healthcare reviews, medical performance factors, text mining, topic modeling, online reviews, online ratings.

Introduction

Online customer reviews have proliferated during the last decade. Prior studies have shown that online reviews significantly influence various essential aspects of a business, such as firm reputation, sales, and product returns (Allard et al., 2020; Ludwig et al., 2013; Sahoo et al., 2018; Zhu & Zhang, 2010). Online reviews have been gaining increasing popularity in the healthcare domain, too (Gao et al., 2015).

A recent survey finds that 72% of patients use online reviews as their first step to finding a new doctor (Hedges & Couey, 2019). Despite the critical role that online healthcare reviews play in patients' provider selection processes, a major concern is that they could be misleading (Gao et al., 2015; Greaves et al., 2012). Unlike online reviews of other goods and services, the stake for online healthcare reviews is high (Lu & Rui, 2018). Biased online reviews may lead to wrong healthcare decisions, which may cause a waste of financial resources and severe damage to provider reputation and patient safety (Brandao et al., 2013; Vanderpool, 2017). Therefore, it is important for both healthcare providers and patients to understand the relationships between objective medical performance measures and online reviews.

Although prior studies have investigated connections between online reviews and actual medical performance, they have produced somewhat mixed and inconsistent results. While some studies find significant results (e.g., Chen & Lee, 2021; Glover et al., 2015), other researchers conclude that there is no significant relationship between online reviews and medical performance (e.g., Saifee et al., 2019; Saifee et al., 2020).

In this study, we attempt to achieve a comprehensive understanding of online healthcare reviews. First, we examine the influence of a representative set of hospital medical performance factors – including readmission, mortality, safety of care, and time in emergency department

(ED) – on online reputation of a hospital. Next, we examine the relationship between a hospital's medical performance and its online reputation from an overall perspective by investigating the effect of *CMS overall quality star rating*, an overall hospital quality measure recently published by the Centers for Medicare and Medicaid Services (CMS).

We also extract an influential factor from review content, *reviewer medical knowledge*, and examine its effect on online reputation. This reflects the medical knowledge level that a reviewer possesses when she posts a review and is measured by the difficulty and frequency of the medical words used in the review. Leveraging the text mining approach, we identify two types of experiential quality evaluations – *service quality evaluations* and *medical quality evaluations* – and unveil the effects of different types of evaluation on online reputation.

To address our research questions, we extract online review data from one of the most popular online review websites, Yelp.com, from 2016 to 2020. Next, we collect hospital data from CMS' *Hospital Compare* system. We then integrate the two datasets by matching the hospitals' unique identification number – National Provider Identifier (NPI) – and the year. To maintain a correct chronological order, we make sure that the hospital measures' end dates precede the review dates. The final dataset consists of 33,582 reviews for 686 hospitals during a period of 5 years, from January 2016 to December 2020. The dataset is at the individual review level so that we can control for reviewers' variations within the same hospital.

Our results indicate that four hospital medical performance factors – readmission, mortality, safety of care, and time in ED – significantly influence online ratings. We also find that reviews written by a knowledgeable reviewer tend to have a lower online review rating. Moreover, we find that the proportion of medical quality evaluations of a hospital in a review enhance its online rating, while more service quality evaluations lead to a lower rating. The significant positive influence of the CMS overall quality star rating on online rating further validates our central thesis that hospitals' medical performance significantly affects their online reputations.

Our findings provide several significant theoretical and practical contributions. First, to the best of our knowledge, this study is the first to empirically examine the effects of various hospital medical performance factors on online reputation within a unified research framework. Specifically, we find that readmission, mortality, safety of care, and time in ED significantly affect online ratings. Prior studies broadly study the effects of online reviews factors on medical performance. Our findings show that hospital medical performance significantly influences a hospital's online reputation. Second, our central thesis that medical performance influences online reputation is supported by the finding that the CMS overall quality star rating significantly affects online rating. CMS publishes this aggregate measure to provide patients with an easy-to-evaluate reference for the overall performance of hospitals. This finding validates the effectiveness of this newly published measure by demonstrating its significant influence on online ratings. Third, we contribute to the literature by introducing an innovative factor in the online healthcare review domain – reviewer medical knowledge – and find that it negatively affects online ratings. Finally, we find the significant effects of two types of experiential quality evaluations – service quality and medical quality evaluations – on online ratings.

The findings of our study have important implications for practice. First, they help us recognize the role played by hospital medical performance factors in shaping a hospital's online reputation. Hospitals could therefore try to improve their online reputation by focusing on and devoting more resources toward improving performance on quality factors such as readmission, mortality, safety of care, and time in ED. Second, the significant influence of CMS overall quality rating shows that it is a reliable indicator for predicting online ratings. Hospitals could rely on the CMS overall quality star rating as a reference to assess their online reputations and take proactive steps that would help in improving perceptions of their performance among online healthcare reviewers.

The findings for reviewer medical knowledge and experiential quality evaluations also have important practical implications for multiple parties. First, although reviews written by knowledgeable reviewers lead to lower online ratings, those reviews are more persuasive (Petty et al.,1981) and reliable (Senecal & Nantel, 2004), thereby having greater impact (Vermeulen & Seegers, 2009). Therefore, it is important for hospitals to interact with the reviewers in more effective ways. For example, hospitals could respond to reviews written by knowledgeable reviewers with low ratings by addressing their concerns and providing explanations. Responding to those reviews can mitigate the influence of low ratings, provide more information, and help improve a hospital's image.

The rest of the essay is organized as follows. We first review the relevant literature in traditional medical performance, online healthcare reviews, and the connections between them. Next, we develop a set of hypotheses. We then describe the data and variables, our empirical analysis, and the results. Finally, we discuss the contributions of our study and conclude the essay by pointing out the limitations and future directions for research.

Literature Review

In this section, we first review relevant research in the areas of hospital performance and online healthcare reviews. Next, we review a set of studies that focus on the relationship between online reviews and hospital medical performance in the healthcare domain.

Traditional Medical Measures

Several institutions have organized programs to publicly report healthcare providers' medical care quality to address information asymmetry in healthcare markets. For example, the physician quality reporting system (PQRS), initiated by CMS, reports on several dimensions of care delivery. Prior studies have shown that clinical outcomes, such as readmission rates and mortality rates, are closely related to patient experience (e.g., Doyle et al., 2013; Glickman et al., 2010). Boulding et al. (2011) find that patient satisfaction is positively correlated with lower readmission rates.

The utility of traditional healthcare report cards has been questioned (Kolstad, 2013; Werner et al., 2012), the relationship between clinical outcomes and patient experience notwithstanding. This is mainly because of their limited propagation among consumers. Brook et al. (2002) find that consumers tend to have difficulty understanding health care report cards because of their professional nature and how they are presented. Meanwhile, online customer reviews for healthcare providers have been steadily growing in popularity (Hanauer et al., 2014; Yaraghi et al., 2018).

Online Healthcare Reviews

As online reviews keep rising in popularity, researchers have started to investigate online reviews in the healthcare domain. Studies examine online healthcare reviews from various angles. Gao et al. (2012) study the relationships between physicians' demographic information and their online ratings. They find that physician specialty, graduation year, certification, medical school ranking, and malpractice claims are significant factors affecting the likelihood of being rated online and the online ratings they receive. Kordzadeh (2019) focuses on the biases of online ratings between official websites and commercial websites. The results indicate systematic differences between the two types of online review websites regarding rating scores and dispersions. Rastegar-Mojarad et al. (2015) build a corpus for patient experience from Yelp reviews.

Another stream of studies examines the influence of online reviews on healthcare providers' demand and patient choice. They commonly find that online ratings positively affect providers' demand (e.g., Luca & Vats, 2013; Segal et al., 2012). Xu et al. (2021) study the influence of review-related variables, such as overall rating and review latent topics, on physicians' 30-day appointments. They find that overall rating has a positive effect on physicians' demand. Moreover, review latent topics increase the predictive power of patient choice by 6% to 12%. The positive relationship between online rating and providers' demand indicates that online healthcare reviews have become a crucial reference source for patients to make decisions (Lu & Rui, 2018). Patients expect providers with largely positive reviews to have better clinical performance (Burkle & Keegan, 2015; Hanauer et al., 2014). Therefore, it is important to investigate if online reviews can be reliable information sources for accurately reflecting healthcare providers' actual medical performance.

Online Reviews and Objective Medical Performance

Prior studies have widely investigated the importance of online reviews in other business domains (e.g., Chevalier & Mayzlin, 2006; Forman et al., 2008; Sahoo et al., 2018). Due to the increasing influence of online reviews in the healthcare domain, a new research stream has emerged, whose focus is to investigate different types of relationships between online reviews and objective medical performance.

Note that this section focuses on the studies that investigate only medical performance factors that reflect healthcare quality of providers, such as readmission and mortality. We do not consider offline survey scores, such as HCAHPS (Ranard et al., 2016), administrative survey score (Daskivich et al., 2018), and consumer checkbook (Gao et al., 2015), as objective measures since they are derived from patients' subjective perceptions in offline settings. Moreover, we do not consider healthcare providers' characteristics, such as physicians' certifications and medical school rankings (Gao et al., 2012), as objective measures since they do not directly reflect medical performance quality. In the rest of this section, therefore, we review papers that investigate at least one objective medical performance factor for healthcare providers.

Studies in this stream typically investigate the relationships between online ratings from various online platforms, such as Yelp.com, Vitals.com, and RateMDs.com, and medical performance factors. Using the difference-in-difference and instrumental variables methods, Chen and Lee (2021) find that online ratings from Yelp are positively associated with traditional clinical measures, including physicians' adherence to clinical guidelines and patients' risk-adjusted health outcomes. Another study finds that lower-rated surgeons are associated with significantly higher in-hospital mortality rates (Lu & Rui, 2018).

In contrast to the findings discussed above, several other studies do not find any significant results (e.g., Saifee et al., 2019). Saifee et al. (2020) study the effects of online review ratings on clinical measures, including readmission rate, ER visit rate, severity level, and mortality risk. However, they do not find any significant results. Similarly, Gray et al. (2015) find that the associations between online physician ratings and clinical quality measures are small and statistically insignificant. Daskivich et al. (2018) examine the impacts of physicians' online ratings on their medical performance, but do not find a significant relationship.

This study aims to extend the findings from prior studies in several important ways. First, most of the extant literature investigates the correlations between healthcare providers' medical performance and their online ratings or examines the impact of online ratings on medical performance (e.g., Glover et al., 2015; Greaves et al., 2014; Daskivich et al., 2018; Lu & Rui, 2018). Bardach et al. (2013), for instance, find that Yelp ratings are positively correlated with lower mortality rates and fewer readmissions. Saifee et al. (2020) use online ratings and review sentiment to predict average clinical performance in the 30 days after the target review date. However, we believe that investigating how medical performance influences online ratings should be a more reasonable direction than the opposite. Consider restaurant reviews on Yelp; a person goes to a restaurant, has a meal, and then writes a review. In the same vein, a patient visits a healthcare provider before writing an online review that provides her perceptions of the quality of service provided during the visit.

Furthermore, although online reputation plays a vital role in patient flow and revenue (e.g., Luca & Vats, 2013; Xu et al., 2021), its impact on actual clinical performance is low. A healthcare provider's medical quality should be determined primarily by its characteristics, such as medical staff quality and equipment level. Therefore, this essay studies the impacts of various medical performance factors on hospitals' online reputation – as reflected in online ratings posted by reviewers.

Second, the selection of medical performance measures has been a heated debate in this research area. Bardach (2018) argues that the medical quality measures adopted by Daskivich et al. (2018) are not representative. There is a lack of an authoritative set of clinical measures that represent the main aspects of healthcare providers' medical performance. In this study, we investigate four independent factors and one composite quality factor, which were recently released by CMS. We believe that the five factors used in this study present a comprehensive and reliable view of the medical quality of healthcare services provided by a hospital. Note that subjective survey scores are often treated as measures of healthcare providers' quality. For instance, HCAHPS is also used for the CMS overall quality rating calculation. However, we exclude it because, like online reviews, HCAHPS surveys are subjective evaluations by patients in offline settings. This study focuses on understanding the relationships between objective quality factors and online reputations for healthcare providers.

Finally, although some studies control for review-related variables, such as review length, latent topics, and readability (e.g., Greaves et al., 2014; Saifee et al., 2020; Xu et al., 2021), none of them controls for reviewer-related variables, such as reviewers' number of friends, reviews, and badges. These variables are significant factors for online reviews in other domains (e.g., Forman et al., 2008; Ghose & Ipeirotis, 2011; Ngo-Ye & Sinha, 2014; Yin et al., 2017; Zhu et al., 2014). Therefore, to account for the differences among reviewers, it is necessary to control for reviewer information in the models for online review ratings. Other than the standard reviewer information directly collected from online platforms, we further investigate three influential review content factors in the online healthcare review domain – *reviewer medical knowledge, medical quality evaluations, and service quality evaluations* – which are extracted from online healthcare review content.

In Table 2.1, we summarize the existing studies in the area and identify the differences between those studies and ours.

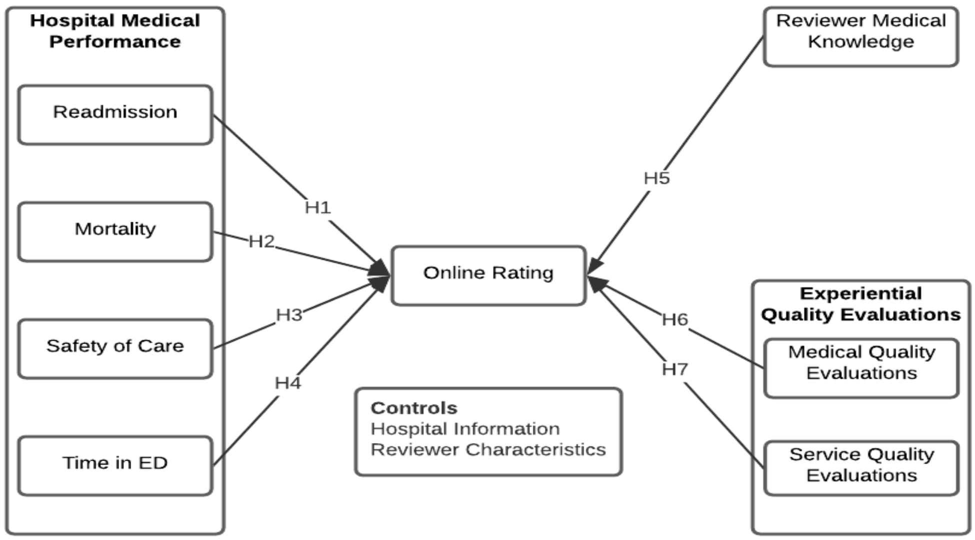
Table 2. 1 Summary of Literature

Study	Outcome	Key Factors	Effects of objective measures on online reputation	Significant objective measures effects	CMS Overall Quality Star Rating	Reviewer Medical Knowledge	Experiential Quality Evaluations
Chen & Lee, 2021	Credentials. Clinical Outcomes. Annual Patient Flow.	Online ratings	No	Yes	No	No	Yes
Gao et al., 2015	Rated online or not. Online rating.	Physician Quality from Consumer's Checkbook	No	Yes	No	No	No
Lu & Rui, 2018	Mortality	Online ratings	No	Yes	No	No	No
Saifee et al., 2019	Online ratings. Review sentiment.	Review content latent topics. Physician Quality Reporting System. EHR use.	Yes	No	No	No	No
Saifee et al., 2020	Discharge rate. Emergency room visits.	Online reviews of physicians	No	No	No	No	No
Xu et al., 2021	30-day physician appointment.	Review latent topics. The number of reviews. Average review ratings. Review readability. Review complexity level. Reviewer's identity.	No	No	No	No	Yes
This Study	Online ratings.	Readmission. Mortality. Safety of care. Time in ED. CMS overall rating. Reviewer medical knowledge. Service-quality evaluations. Medical quality evaluations.	Yes	Yes	Yes	Yes	Yes

Hypothesis Development

In this section, we develop our research hypotheses. Figure 2.1 presents our research model.

Figure 2. 1 Research Model



Hospital Medical Performance and Online Reputation

The relationship between objective quality and perceived quality has been investigated for a long time. By definition, objective quality is the aggregate performance of all product attributes, which generally does not include intangible attributes such as aesthetics and extrinsic attributes (Mitra & Golder, 2006). Prior studies commonly measure objective quality using composite instruments and expert ratings (e.g., Curry & Riesz, 1988; Lichtenstein & Burton, 1989; Riesz, 1978). On the other hand, perceived quality is defined as the overall subjective judgment of quality relative to the expectation of quality. These expectations are based on one's own and others' experiences, as well as various other sources, including brand reputation, price, and advertising (Boulding et al., 1993; Johnson et al., 1995; Zeithaml, 1988).

Previous studies find a significant positive effect of objective quality on perceived quality (Bolton & Drew, 1991b; Landon & Smith, 2012). Three different studies consistently find that subjective quality positively reflects 9% to 22% of the change in objective quality (Boulding et al., 1993; Boulding et al., 1999; Prabhu & Tellis, 2000). Similarly, two field studies also find positive and significant relationships between objective quality and subjective quality (Bolton & Drew, 1991a; Kamakura et al., 2002). Mitra and Golder (2006) find that 20% of the change in objective quality can be reflected in customer perceptions of quality in the first year.

The objective quality of a hospital can be assessed from various perspectives. In this study, we test the effects of four hospital medical performance factors – readmission, mortality, safety of care, and time in ED – on online rating.

Readmission

In our study, we measure readmission with a single composite measure – the 30-day hospital-wide readmission rate. The hospital-wide all-cause readmission measure is included in the Hospital Inpatient Quality Reporting (IQR) Program. This measure estimates a risk-standardized readmission rate of unplanned, all-cause readmission within 30 days of hospital discharge (CDP, 2019; Horwitz et al., 2011). It is "a claims-based, risk adjusted hospital-wide readmission (HWR) measure for public reporting that reflects the quality of care for hospitalized patients in the United States" (Horwitz et al., 2011).

Several studies find an association between quality of inpatient care and early readmission rates for a wide range of conditions. Prior studies have shown that it is possible for hospitals to reduce readmission rates through quality-of-care initiatives. Readmission rate is an important measure of the actual quality of hospital care (Salinas, 2017). It is negatively associated with quality culture of the hospital (Lee et al., 2018), which emphasizes efficiency of care operations, standardization of care processes, higher dedication among physicians, etc. We can therefore treat readmission rate as an effective quality measure. As discussed above, prior studies have found a significant positive influence of objective quality on perceived quality (Bolton & Drew, 1991b; Kamakura et al., 2002; Landon & Smith, 2012). Based on this, we argue that the readmission rate, which is an objective quality measure, will have a significant impact on the perceptions of a hospital's overall quality by patients, as expressed through online ratings they assign on online review platforms. Hence, we state the following hypothesis:

H1: Hospitals with higher readmission rates receive lower online ratings.

Mortality

The mortality is measured by averaging the 30-day death rates from CMS data for the six conditions and procedures: heart attack, coronary artery bypass graft (CABG), chronic obstructive pulmonary disease (COPD), heart failure, pneumonia, and stroke. Kassirer (1999) points out that "… there is a general assumption that the teaching hospitals provide better care than nonteaching hospitals. They have a greater concentration of clinical expertise, a focus on clinical research, and technological superiority. They also score better in the national analysis of the quality of hospital care performed each year by the respected National Opinion Research Center at the University of Chicago …." (p. 309). Allison et al. (2000) demonstrate that mortality for patients treated at non-teaching hospitals is greater than that for those treated at major teaching hospitals. This has also been observed by Burke et al. (2017). Based on the finding that higher mortality rates are associated with lower quality hospitals, we argue that mortality rate exerts a negative influence on the online perceptions of a hospital's quality. We therefore propose the following hypothesis:

H2: Hospitals with higher mortality rates receive lower online ratings.

Safety of care

We use a single composite indicator from the Hospital Compare site, PSI_90_SAFETY, to measure safety of care. The Agency for Healthcare Research and Quality (AHRQ) developed 26 Patient Safety Indicators (PSIs) that health providers can use to identify potential in-hospital patient safety problems. The CMS Patient Safety and Adverse Events Composite (PSI_90_SAFETY) is a composite measure that summarizes patient safety problems across multiple indicators, monitors performance over time, and facilitates comparative reporting and quality improvement at the hospital level. A higher PSI_90_SAFETY score reflects a worse performance with respect to patient safety during the delivery of care.

Prior studies have found an association between quality of care and problems relating to safety of care. The patient safety score represents a set of measures on potential complications and adverse events that hospitalized patients could experience (Bonis et al., 2008). Initiatives for improving safety and quality of care should include targeting patients with postoperative complications (Khan et al., 2006). For example, Geraci et al. (1999) find that better quality of care is associated with a lower incidence of in-hospital complications in patients who are hospitalized for diabetes or COPD. Bergman et al. (2014) also find that a higher quality score is associated with fewer postoperative complications. Based on these findings, we argue that higher incidence of safety problems signals a lower quality score for a hospital. Hence, we state the following hypothesis:

H3: Hospitals where patients experience a higher incidence of safety problems receive lower online ratings.

Time in ED

The time in emergency department measures belong to the quality dimension of timely and effective care. We only consider the measures for the time in ED because they are more common and representative than other measures in the timely and effective care dimension. Other measures are specialized for specific conditions such as chest pain, stroke symptoms, and severe sepsis, or for medical tests such as colonoscopy, radiation therapy, and MRI. These conditions and tests are too limited to represent the medical quality of an entire hospital.

We use two Hospital Care measures for time in ED: ED_2b and OP_18b. ED_2b refers to the median time a patient spends in the emergency department after the doctor decides to admit her as an inpatient. OP_18b is the median time a patient spends in the emergency department before being sent home. Since both these measures represent the time spent in ED, we take their average to measure the Time in ED variable. Prior studies have shown that the time patients spend in ED is an indicator of a hospital's quality of care. Longer Emergency Department-Length of Stay (ED-LOS) has been used as a proxy for ED crowding (Andersson et al., 2020). Mullins and Pines (2014) find that worse performance on ED crowding is associated with lower patient satisfaction. Montes et al. (2019) find that prolonged ED-LOS is associated with increased adverse "perioperative outcome" for patients. Given the adverse effects associated with longer ED-LOS, we state the following hypothesis:

H4: Hospitals where patients spend more time in ED receive lower online ratings.

Reviewer Medical Knowledge and Online Reputation

In this study, reviewer medical knowledge refers to the level of medical knowledge that a reviewer possesses when she writes the review, as reflected in the online review content. In general, the effects of consumer expertise on product evaluations have been widely studied for a long time (e.g., Bettman & Park, 1980; Cordell, 1997; Johnson & Russo, 1984; Maheswaran et al., 1996; Park & Lessig, 1981). More recently, examining the effects of consumer expertise has been gaining popularity in online business domains (e.g., Kim et al., 2011; Zou et al., 2011).

Several studies find that reviewers' expertise negatively influences online reputation (e.g., Lawrence & Perrigot, 2015). Guo and Zhou (2016) argue that experts tend to write reviews that contain negative information to maintain influence and attract attention. Therefore, experts tend to give lower review ratings and write more negative content in review text than novices (Yin et al., 2014). Han (2021) finds that a high-expertise reviewer gives a lower rating and contains more negative content than a low-expertise reviewer. The negative relationship between reviewer expertise and online reputation could be because of high expectations. Researchers argue that high-expertise customers tend to have higher expectations (Reinartz & Kumar, 2002) and more restricted evaluation criteria (Ladhari et al., 2011).

According to the conceptual model of service quality (Parasuraman et al., 1985), perceived service quality is a comparison between expected service and perceived service. When expected service exceeds perceived service, perceived quality becomes negative. Therefore, if a reviewer expresses a high level of medical knowledge in her review, she would have a higher expectation of her hospital visit and, therefore, evaluate the hospital with a set of stricter criteria. As a result, her online review ratings would tend to be lower than the ratings given by those with a lower level of medical knowledge. We therefore pose the following hypothesis:

H5: Reviewer medical knowledge has a negative influence on a hospital's online rating.

Experiential Quality Evaluations

According to the expectation-confirmation theory (ECT) (Oliver, 1980), consumers' product evaluation is a function of experiential quality and expectation, where the latter is formed before making purchase decisions. Experiential quality is derived from consumers' experience with products or services. In this study, we extract evaluations of experiential quality from the content of online reviews, where reviewers share their hospital visit experience. Online ratings are consumers' overall product evaluations provided as part of their online reviews. Although writing a review and giving a star rating happen simultaneously in online platforms, experiential quality evaluations help generate the overall evaluation (Oliver, 1980). Therefore, we investigate the effects of two experiential quality evaluations embedded in the review contents – service quality evaluation and medical quality evaluation – on a hospital's online reputation.

Because of the limited medical knowledge that patients possess and the complexity of healthcare services, they are likely to have low expectations regarding the medical performance of the hospitals they visited. As a result, they are more likely to be satisfied by the medical quality of the hospitals. Patients come to hospitals for some health conditions. As long as their conditions get healed or relieved, they will feel satisfied, even if the treatments are not optimal and there exist some flaws in the processes; recognizing these flaws tends to require a high level of medical knowledge, which most patients do not possess. To assess the literature on the concept and measurement of patients' expectations for healthcare and develop a measure of patients' expectations, Bowling et al. (2012) review 211 papers, search five major electronic databases, and survey 833 patients before finding that among the 27 items of pre-visit expectations, patients have the five lowest ideal expectations related to clinical procedures, including physical examination, tests/investigations, diagnosis, prescription, and referral.

On the other hand, patients usually have high expectations with respect to the service quality of hospitals, since they have richer prior experience from other services, such as hotels and restaurants. The majority of the highest ideal expectations that Bowling et al. (2012) identify are related to service quality, such as cleanliness, information about where to go, having convenient appointments, being seen on time, helpfulness of reception staff, and having a clear and easy to understand doctor.

Previous research argues that prior experience is a significant determinant of consumer satisfaction via its impact on expectations (LaTour & Peat, 1979; Woodruff et al., 1983). Satisfaction is not absolute but relative to one's past experience (LaTour & Peat, 1980). Although patients may have limited experience with hospitals, they can accumulate their experiences of services from various other service providers. Therefore, when they visit a hospital, they may use their prior experience from other services as a basis for evaluating a hospital's service quality. However, the healthcare domain is essentially different from other businesses in terms of their priorities. Unlike other businesses commonly treating service quality as one of their top priorities, healthcare providers' top priorities focus on saving lives and healing patients. Therefore, if patients evaluate services provided by hospitals based on their prior experience in other business domains, they are likely to generate dissatisfaction and express negative feelings through online reviews.

Overall, we believe that patients' evaluations of different hospital experiential qualities have different effects on their online reputation. It is because patients tend to have different levels of expectations for the medical quality and the service quality at a hospital. Because people commonly have limited medical knowledge and prior medical experience, they are more likely to have low expectations of a hospital's medical performance. In contrast, people usually have a rich experience with services, which they develop through past experiences in other domains. Therefore, they would expect to receive the same level of service from a hospital as that of the services they received from other businesses. As a result, they are more easily disappointed by a hospital's services.

Moreover, people are more willing to share their experiences when it exceeds or fails their expectations. This is apparent in our data; about 83% of the reviews have a rating of 1 or 5 on a 5-point scale. Therefore, reviewers share their medical experiences because most likely they exceed their expectations. On the contrary, they share their service experiences because most likely they were disappointed. Hence, we propose the following hypotheses:

H6: A review that contains more evaluations on the medical quality of a hospital is likely to yield a higher online rating.

H7: A review that contains more evaluations on the service quality of a hospital is likely to yield a lower online rating.

Data

The data set we use spans five years, from January 2016 to December 2020. The hospital information is obtained from the *Hospital Compare* system, covering measures that reflect various aspects of a hospital's medical performance. Hospital Compare includes information on more than 100 quality measures for over 4,000 acute care and critical access hospitals nationwide. With the exception of the Patient Experience group – which includes measures based on a patient's experience at a hospital – the measures in all groups rely on actual hospital performance data (e.g., readmission rates, mortality rates, safety problems, etc.). Different measures are stored in different tables, such as "hospital general information," "complication and death," "timely and effective care," and "unplanned hospital visits" by year. We first integrate hospital measures from different tables into a single table by year. Because the participation in the *Hospital Compare* program is voluntary[3], the participating hospitals and available measures vary over the years. Therefore, when we merge the tables from 2016 to 2020 into one dataset, we only reserve the hospitals that are available over all five years.

Moreover, since hospitals are not required to report all the measures, some of them are largely incomplete. To ensure the reliability of the measures, we further reduce the integrated dataset by removing the hospitals that miss more than two years' key measures used in our models, including the hospital-wide 30-day readmission rate (READM_30_HOSP_WIDE), the six mortality rates (MORT_30_AMI, MORT_30_CABG, MORT_30_COPD, MORT_30_HF, MORT_30_PN, MORT_30_STK), the composite patient safety indicator (PSI_90_SAFETY), and two timely and effective care in ED care (ED_2b, OP_18b). The other missing values are replaced by the corresponding measures' average values in the years. The final integrated dataset contains 686 hospitals across five years. The 686 hospitals are in 49 states across the US.

After finalizing the hospital list, we collect their review information, which mainly contains review-related and reviewer-related information, from Yelp.com. Yelp is an online platform where users can share their evaluations online (ratings and textual reviews) of businesses such as restaurants, hotels, and hospitals. It is one of the most widely adopted commercial online review website in the United States for hospital ratings (Bardach et al., 2013). We ensure the match of hospital and review information by matching unique hospital National Provider Identifier (NPI) numbers. We extract 34,130 reviews for the 686 hospitals over five years. The number of reviews for each hospital ranges from 2 to 392. We further eliminate

[3] https://qualitynet.cms.gov/inpatient/public-reporting/public-reporting/participation

reviews with less than six words since concise reviews tend to contain little information. Finally, our integrated dataset contains 33,582 reviews from 696 hospitals across five years.

Different hospital measures from CMS may have different end dates. We therefore check the end date for each measure and ensure that it is earlier than the corresponding review date. We match the hospital and review information by year. For instance, for all the reviews posted in 2018, we collect values of hospital performance measures on or before 12/31/2017. Maintaining an appropriate chronological order allows us to examine the causal effects of hospital performance factors on online reputation.

Variables

Dependent Variable

The dependent variable is online rating. It is the rating that a reviewer gives to a hospital on a 5-point scale on Yelp.com, indicating her overall evaluation of the hospital.

Hospital Medical Performance Variables

In this essay, we are interested in examining the effects of the hospital medical performance factors on online rating. We extract the hospital medical performance measures from the Hospital Compare datasets. As discussed before, we employ four medical performance variables: hospital wide readmission rate, average mortality rate, composite patient safety indicator, and average time in ED.

Reviewer Medical Knowledge

We extract reviewers' medical knowledge level at the time they write their reviews from online review contents by matching those contents with the consumer health vocabulary (CHV). CHV is designed to complement existing knowledge in the Unified Medical Language System (UMLS). Its terms focus on expressions and concepts that are employed by health-related communications from or to consumers. The vocabulary provides several indicators that reflect the frequency and difficulty of the terms, including frequency scores, context scores, CUI scores (difficulty of the concept), and combo scores. The combo score is a composite measure that combines the other three measures. Combo scores range from 0 to 1; the lower the score, the more difficult and less frequent the term is.

We measure the medical knowledge of a reviewer by averaging the combo scores for all the matched terms in a review. We believe that a combo score reflects reviewer medical knowledge since only reviewers who have enough medical knowledge can use those complex medical terms to describe their experiences.

The following two reviews are examples from our data. They both describe the symptoms and comments on doctors and nurses.

Review A: "*My wife had A-Fib arrhythmia and irregular heartbeats…. The doctor dropped her from the IV and double her pill dose to 12.5 mg and left…. However, she was already taking 12.5 mg. 2x daily for the past 8 years….*"

Review B: "*I have had bad pressure on my head, at times I would see flashing lights!!... The nurse or whoever he was I asked for a blanket because I was cold he said rudely yeah and so when he did come back in he threw the blanket on me I said thanks he said yep !!...*"

Although both reviews describe the symptoms and provide evaluations of the medical staff, review A uses more professional and complex terms than review B. This is reflected by their combo scores: review A's score is 0.277, review B's score is 0.5. Note that the combo score is a reverse measure. A lower combo score represents higher medical knowledge expressed in the review.

Experiential Quality Evaluations

We exploit reviewers' evaluations of medical and service quality from review content. We extract latent topics from review content by applying the Topic Modeling technique. One of the most popular topic-modeling approaches is Latent Dirichlet allocation (LDA) (Blei et al., 2003). LDA is an unsupervised method that exploits a predefined number of hidden topics from documents. Topics are groups of terms that tend to co-occur. Those groups can then be interpreted and labeled by human coders based on their content. We conduct the topic modeling by using the LDA implementation in MALLET (Machine Learning for Language Toolkit) (McCallum, 2002) via a Python package called Gensim (Rehurek & Sojka, 2011).

First, we build 25 models with the number of topics ranging from 2 to 50, and with a step equal to 2. The results show that the model with 20 topics yields the highest coherence score. Therefore, we build an LDA model for 20 topics. The outputs are topics' keyword lists and their proportions to reviews. The topic modeling approach assumes that each document contains all topics and is only different in terms of proportions. Therefore, the total contribution of the 20 topics is 1.

Then, according to the corresponding top 10 keywords, we give a name to each of the 20 topics. The keyword lists and topic names are presented in Table 2.2. For example, Topic 1 has keywords such as "doctor," "test," "diagnosis," and "lab." They are related to a doctor's diagnosis process. Thus, we name Topic 1 "Diagnosis." Topic 2 corresponds to the insurance and billing with terms, such as "bill," "insurance," "charge," and "money." So we name it "Insurance and Billing." Topic 3 discusses various pains. Thus, we name it "Pain Control." Topic 4 contains keywords that are about discharge information and back-home care plan. Thus, we name it "Discharge Info." Notably, seven topics contain mixed and less informative keywords. Therefore, we name them “Other.” Although prior studies investigate the effects of review latent topics (e.g., Greaves et al., 2014; Saifee et al., 2019), only a few recent studies further summarize topics into higher classification categories. Following Chen and Lee (2021), we next classify the latent topics into three groups: “service quality evaluations,” “medical quality evaluations,” and “other.”

As shown in Table 2.2, in general, we identify six “medical quality evaluations” topics, seven “service quality evaluations” topics, and seven “other” topics. Specifically, the topics “Diagnosis,” “Pain Control,” “Treatment,” “Surgery,” “Baby Delivery,” and “Family Care” are classified as “medical quality evaluations.” It is because these topics are related to medical procedures or treatments. Topics “Insurance and Billing,” “Discharge Info,” “Responsiveness,” “Nurse Communication,” “Room Cleanliness,” “Food,” and “Waiting Time” are recognized as “service quality evaluations” since they describe service aspects of a hospital. The remaining seven “Other” topics are grouped into “other”.

Finally, we sum the proportions of all the topics that belong to the same group as the measures. The two focal variables – medical quality evaluation and service quality evaluation – are measured as the total proportions of their corresponding topics.

Table 2. 2 Topic Information

Topic	Keywords	Topic Name	Group
1	doctor, test, result, order, visit, treatment, problem, diagnosis, show, lab	Diagnosis	Medical Quality
2	bill, pay, insurance, charge, receive, send, service, money, cost, visit	Insurance and Billing	Service Quality
3	pain, back, arm, ray, leave, fall, break, leg, hurt, home	Pain Control	Medical Quality
4	back, home, check, send, leave, discharge, bring, start, sick, fine	Discharge Info	Service Quality
5	call, appointment, time, phone, answer, back, speak, information, talk, number	Responsiveness	Service Quality
6	nurse, time, feel, question, check, experience, explain, understand, make sure, talk	Nurse Communication	Service Quality
7	medical, facility, review, year, health, physician, system, case, include, resident	Other	Other
8	place, doctor, time, emergency, run, avoid, joke, awful, sick, top	Other	Other
9	visit, area, friend, time, close, parking, small, facility, staff, walk	Other	Other
10	room, nurse, bed, floor, night, clean, move, bathroom, dirty, chair	Room Cleanliness	Service Quality
11	patient, treat, care, treatment, employee, respect, compassion, understand, attitude, poor	Treatment	Medical Quality
12	walk, leave, back, talk, hand, hear, put, door, start, head	Other	Other
13	care, food, stay, staff, cafeteria, receive, meal, service, eat, lunch	Food	Service Quality
14	staff, care, experience, service, quick, pleasant, positive, fast, recommend, visit	Other	Other
15	surgery, day, husband, procedure, wife, experience, surgeon, time, recovery, perform	Surgery	Medical Quality
16	hour, wait, minute, sit, time, waiting room, long, check, triage, back	Waiting Time	Service Quality
17	refuse, state, write, medication, lie, sign, wrong, report, request, happen	Other	Other
18	life, die, day, heart, save, week, year, infection, due, lose	Other	Other
19	nurse, baby, daughter, experience, son, child, deliver, time, delivery, room	Baby Delivery	Medical Quality
20	nurse, family, mom, mother, care, day, admit, dad, family member, transfer	Family Care	Medical Quality

Control Variables

Prior online review studies have largely recognized the importance of reviewers' characteristics for search or experience products or services. For example, by examining the restaurant reviews, Zhang and Liu (2019) find that reviewers' number of friends negatively affects the review ratings. However, as one of the most typical credence services, the impacts of reviewers' characteristics on healthcare reviews have been overlooked. Following the previous online review literature, we control for reviewer-related variables, including reviewer number of friends (Banerjee et al., 2017), number of reviews (Cheng & Ho, 2015), number of badges (Zhu et al., 2014), and number of followers (Luo et al., 2021). In addition to reviewer information, review and review content information have consistently been recognized as important factors in the online review domain (Ghose & Ipeirotis, 2010; Yin et al., 2014). Therefore, we further control for review length, subjectivity, and Fog index.

We also use the following hospital control variables in our models: number of beds, emergency department volume, hospital type, hospital ownership, and hospital state. Table 2.3 summarizes the information for all the variables used in this essay.

Table 2. 3 Variable Summary

Group	Variable Name	Description	Summary Statistics: Mean (S.D.)
Dependent Variables	Online Rating	The online rating rated by reviewers for hospitals in Yelp.com.	2.606 (1.826)
Focal Variables: Objective Medical Performance Factors	CMS Overall Rating	The CMS overall quality star rating.	2.885 (1.097)
	Safety of Care	The CMS Patient Safety and Adverse Events Composite (CMS PSI 90).	1.303 (0.629)
	Readmission	The hospital-wide rate of 30-days readmission after discharge from hospitals.	15.511 (0.962)
	Mortality	The average of death rates for six different diseases: heart attack, CABG surgery, COPD patients, heart failure, pneumonia, and stroke	12.146 (1.159)
	Time in ED	The waiting time in emergency departments. The average of ED_2b and OP_18b.	156.743 (48.734)
Focal Variables: Review Content Factors	Reviewer Medical Knowledge	The average combo score of the terms that match the consumer health vocabulary (CHV). It ranges from 0 to 1. The higher the score easier the term is.	0.343 (0.094)
	Medical Quality Evaluations	The total proportions of latent review topics that evaluate the medical quality of hospitals.	0.298 (0.047)
	Service Quality Evaluations	The total proportions of latent review topics that evaluate service quality of hospitals.	0.353 (0.052)
Control Variables: Reviewer Information	Reviewer Number of Friends	The number of friends of a reviewer.	64.961 (230.356)
	Reviewer Number of Reviews	The number of past reviews posted by a reviewer.	54.772 (275.734)
	Reviewer Number of Badges	The number of badges a reviewer has received.	0.356 (1.399)
	Reviewer Number of Followers	The number of followers of a reviewer.	2.936 (33.763)
Control Variables: Review Information	Review Length	The number of words of a review.	168.638 (161.668)
	Review Subjectivity	Subjectivity score of a review ([0,1]=[objective, subjective])	0.528 (0.154)
	Review Fog Index	Fog index of a review	15.113 (18.216)
Control Variables: Hospital Information	Number of Beds	Hospital number of beds.	392.64 (348.39)
	EDV	Hospital emergency department volume.	Categorical Variable
	Hospital Type	Hospital type.	Categorical Variable
	Hospital Ownership	Hospital ownership.	Categorical Variable
	Hospital State	The state where the hospital is located.	Categorical Variable

Data Analysis

Impact of Hospital Medical Performance on Online Perceptions

We test the effects of a set of hospital performance variables, including readmission, mortality, safety of care, and time in ED, on online rating using the model shown below:

$$\begin{aligned} Online\ Rating_{rjt} &= \beta_0 + \beta_1 Readmission_{rj(t^-)} + \beta_2 \mathrm{Mortality_{CMS_{rj(\mathit{t}^-)}}} + \beta_3 \mathrm{SafetyofCare_{rj(\mathit{t}^-)}} \\ &+ \beta_4 \mathrm{Time\ in\ ED_{rj(\mathit{t}^-)}} + \beta_5 \mathrm{ReviewerMedicalKnowledge_{rjt}} \\ &+ \beta_6 \mathrm{MedicalQualityEvaluations_{rjt}} + \beta_7 ServiceQualityEvaluations_{rjt} \\ &+ \beta_8 NumberofFriend\_Reviewer_{rjt} + \beta_9 NumberofReviews\_Reviewer_{rjt} \\ &+ \beta_{10} NumberofBadges\_Reviewer_{rjt} + \beta_{11} \mathrm{NumberofFollowers\_Reviewer_{rjt}} \\ &+ \beta_{12} ReviewLength_{rjt} + \beta_{13} Subjectivity_{rjt} + \beta_{14} FogIndex_{rjt} \\ &+ \beta_{15} NumberofBeds_{rjt} + \beta_{16} EDV_{rjt} + \beta_{17} HospitalType_{rjt} + \alpha_{\mathrm{j}} + \gamma_{\mathrm{j}} \\ &+ \varepsilon_{rj} \qquad (1) \end{aligned}$$

The dependent variable of equation (1) is online rating. β_1 to β_4 measure the effects of readmission, mortality, safety of care, and time in ED, respectively, for hospital *j* reviewed by reviewer *r* with the end dates in year *t-1*. Note that, although the end dates are in year *t-1*, some measures last for more than one year. For example, the safety of care measure, CMS PSI 90, in 2020 is measured from 7/1/2017 to 6/30/2019. Therefore, in the equation (1), t^- represents the time before year *t*. β_5 captures the effect of reviewer medical knowledge expressed in review *r* for hospital j in year *t*. β_6 and β_7 represent the effects of medical quality evaluations and service quality evaluations, respectively. β_8 to β_{17} are coefficients of the control variables. α_{j} denotes the fixed effect of hospital ownership, γ_{j} represent the fixed effect of hospital state, and ε_{rj} represents the error term.

Omitted Variable Bias

The hospital performance measures may be biased because of omitted variables. Omission of variables can violate the exogeneity assumption if the omitted variable associated with the dependent variable is also correlated with the independent variables (Kennedy, 2008; Wooldridge, 2002), for example, when patients have a selection bias. Patients with more severe conditions tend to seek higher-quality hospitals. Meanwhile, more severe conditions lead to a higher probability of failures or problems, such as safety problems, readmissions, and even deaths. Those harmful consequences may cause patients to have wrong impressions of hospitals and, as a result, leave negative reviews online. In general, patients' selection bias may be correlated with both the dependent variable (online rating) and the independent variables (hospital medical performance). Moreover, the delivered medical quality can be different for different hospital visits, which may also bias the results. The same hospital may deliver different levels of quality depending on various factors, such as physicians, nurses, and patient volume.

To address this concern, we employ the coefficient stability approach proposed by Oster (2019). This approach argues that the robustness of estimates to omitted variable bias can be examined by observing movements in the coefficient of the focal explanatory variable (also called the treatment) and the R-squared from a baseline model that only includes the treatment to a model consisting of a complete set of control variables. If the coefficient of the treatment only changes slightly from the baseline model with a small R-squared to the full model that has a substantial increase in the R-squared, it indicates that the estimate is robust. To set the maximum R-squared (R_{max}), Oster argues that it is reasonable to set the value as 1.3 times the full model's R-squared. Our empirical analyses report the adjusted coefficient using the coefficient stability method proposed by Oster (2019).

Empirical Results

Model 1 in Table 2.4 shows the results of the four hospital medical performance variables with reviewer and hospital information controlled. The correlation matrix in Appendix I shows that the correlations among the focal medical performance and review content variables are reasonably low. Some reviewer information variables have relatively high correlations. However, we inspect the variance inflation factor (VIF) for each model, and find that the VIFs of all independent variables are well below the generally established threshold of 10 (Hair et al., 1998), indicating that multicollinearity is unlikely to confound our findings. As the results show, readmission ($\alpha = -0.02448, p < 0.01$), mortality ($\alpha = -0.03407, p < 0.001$), safety of care ($\alpha = -0.09348, p < 0.001$), and time in ED ($\alpha = -0.00315, p < 0.001$) have significant negative effects on online ratings, which support hypotheses H1, H2, H3, and H4. The coefficient stability results for the four medical performance variables show that their adjusted coefficients are also negative; thus, it proves that the omitted variables do not bias the estimation results.

Impact of Reviewer Medical Knowledge and Experiential Quality Evaluations

Other than the hospital medical performance factors, we are also interested in the effects of review content factors on online perceptions. This study investigates three innovative review content factors in the online healthcare review area: reviewer medical knowledge, medical quality evaluations, and service quality evaluations. Reviewer medical knowledge is measured by the average difficulty and frequency of the medical terms used in a review.

Compared to Model 1 in Table 2.4, Model 2 includes these three review content factors. Reviewer medical knowledge has a significant positive coefficient ($\alpha = 0.4126, p < 0.001$). It is a reverse measure, which means that the lower the value, the higher is the medical knowledge of the reviewer. Therefore, the positive coefficient means that a knowledgeable reviewer tends to give a lower rating, thereby supporting hypothesis H5. As for experiential quality evaluations, medical quality evaluation has a positive significant coefficient ($\alpha = 1.837, p < 0.001$), whereas service quality evaluation ($\alpha = -2.245, p < 0.001$) negatively affects online perceptions. The results support hypotheses H6 and H7. They indicate that if a review focuses on evaluating a hospital's medical quality, it is more likely to assign a higher online rating to the hospital. Note that the results for the four hospital performance factors are consistent with the results in Model 1, which further confirm the findings for those factors. Moreover, the higher R-square value shows that including the three review content factors improves model fit.

Table 2. 4 Empirical Estimation and Coefficient Stability Results

Variables	Model 1	Model 2
(Intercept)	3.613 (0.4619) ***	3.849 (0.476) ***
Readmission	-0.02448 (0.00843) **	-0.02526 (0.00838) **
Mortality	-0.03407 (0.00671) ***	-0.03414 (0.00667) ***
Safety of Care	-0.09348 (0.0162) ***	-0.09327 (0.0161) ***
Time in ED	-0.00315 (0.00022) ***	-0.00314 (0.00022) ***
Reviewer Medical Knowledge		0.4126 (0.09812) ***
Medical Quality Evaluations		1.837 (0.2301) ***
Service Quality Evaluations		-2.245 (0.204) ***
Reviewer Number of Friends	0.00022 (0.00005) ***	0.00021 (0.00005) ***
Reviewer Number of Reviews	0.00007 (0.00005)	0.00007 (0.00005)
Reviewer Number of Badges	0.2113 (0.00766) ***	0.2126 (0.00763) ***
Reviewer Number of Followers	-0.00097 (0.00045) *	-0.00102 (0.00045) *
Review Length	-0.00189 (0.00006) ***	-0.00192 (0.00006) ***
Review Subjectivity	1.694 (0.0619) ***	1.643 (0.06157) ***
Review Fog Index	-0.00105 (0.0005) *	-0.00089 (0.00049)
Hospital Number of Beds	0.00005 (0.00003)	0.00006 (0.00003)
EDV: Medium	0.1126 (0.03512) **	0.104 (0.03492) **
EDV: High	0.03121 (0.0374)	0.02549 (0.03718)
EDV: Very High	0.02098 (0.03641)	0.01901 (0.0362)
Hospital Type: Critical Access Hospitals	0.225 (0.1006) *	0.2094 (0.09999) *
Hospital Ownership	Yes	Yes
Hospital State	Yes	Yes
R Square	0.1097	0.1206

*** p<0.001, ** p<0.01, * p<0.05

Coefficient Stability	Readmission	Mortality	Safety of Care	Time in ED
Adjusted Beta	-0.0315	-0.0326	-0.0896	-0.00302
Controlled Coefficient	-0.036	-0.034	-0.103	-0.00309
Controlled R Square	0.116	0.116	0.117	0.121
Max R Square	0.151	0.151	0.152	0.157

Robustness Check: Impact of CMS Overall Quality Star Rating

In the main models, we show that multiple hospital medical performance factors, including readmission, mortality, safety of care, and time in ED, significantly affect hospitals' online ratings. In this section, we conduct a robustness test using a measure that captures a hospital's overall performance.

In July 2016, CMS first released a composite measure of hospital quality consolidated from various measures across five quality areas, including mortality, safety of care, readmission, patient experience, and timely and effective care, into a familiar 5-star rating system. This summary score is calculated by taking the weighted average over the five quality areas. The score is then used to assign hospitals to different star ratings by k-means clustering[4]. The measures that we have used to represent the four hospital medical performance factors in prior sections are all used in the calculation of the CMS overall quality star rating.

Despite some criticisms (Bilimoria & Barnard, 2016, 2021), many studies find that the CMS overall quality star rating has significant correlations with various medical outcomes for an extensive range of treatments, such as arthroplasty, coronary artery bypass grafting (Fontana et al., 2019), advanced laparoscopic abdominal surgery (Koh et al., 2017), and complex cancer surgery (Mehta et al., 2020; Papageorge et al., 2020). However, none of the prior studies examines the relationships between the CMS overall quality star rating of a hospital and its online reputation.

In this section, we examine the influence of the CMS overall quality star rating on online ratings. We use the same set of control variables as in the main models (see Table 2.3).

[4] More information about the overall hospital quality star rating can be found on the CMS website: https://data.cms.gov/provider-data/topics/hospitals/overall-hospital-quality-star-rating/.

Omitted Variable Bias

Previously, we handled potential omitted variable bias for the four hospital medical performance variables by using the coefficient stability method (Oster, 2019). Here, we apply the same method for the CMS overall quality star rating for addressing the omitted variable issue.

Empirical Results

Model 1 in Table 2.5 show the results of the model for the CMS overall quality star rating, and Model 2 further includes the review content factors. The significant positive coefficient of the CMS overall quality star rating ($\alpha = 0.0955, p < 0.001$) in Model 1 indicates that a hospital's overall quality significantly influences its online reputation. The results provide further evidence for our thesis that a hospital's medical performance quality significantly influences its online reputation. Also, the effects of the three review content factors are consistent with the results in the main model.

Table 2. 5 Empirical Estimation and Coefficient Stability Results II

Variables	Model 1	Model 2
(Intercept)	1.81 (0.4308) ***	2.023 (0.4461) ***
CMS Overall Quality Star Rating	0.0955 (0.00942) ***	0.09413 (0.00936) ***
Reviewer Medical Knowledge		0.4126 (0.09834) ***
Medical Quality Evaluations		1.863 (0.2306) ***
Service Quality Evaluations		-2.214 (0.2045) ***
Reviewer Number of Friends	0.00023 (0.00005) ***	0.00022 (0.00005) ***
Reviewer Number of Reviews	0.00008 (0.00005)	0.00008 (0.00005)
Reviewer Number of Badges	0.212 (0.00768) ***	0.2133 (0.00764) ***
Reviewer Number of Followers	-0.00108 (0.00046) *	-0.00114 (0.00045) *
Review Length	-0.0019 (0.00006) ***	-0.00193 (0.00006) ***
Review Subjectivity	1.701 (0.06203) ***	1.65 (0.06171) ***
Review Fog Index	-0.00103 (0.0005) *	-0.00086 (0.0005)
Hospital Number of Beds	-0.00006 (0.00003)	-0.00005 (0.00003)
EDV: High	-0.06444 (0.037)	-0.07011 (0.03678)
EDV: Medium	0.08539 (0.03509) *	0.0766 (0.03488) *
EDV: Very High	-0.07226 (0.03577) *	-0.07449 (0.03556) *
Hospital Types: Critical Access Hospitals	0.2498 (0.1008) *	0.2348 (0.1002) *
Hospital Ownerships	Yes	Yes
Hospital States	Yes	Yes
R Square	0.1055	0.1164

*** p<0.001, ** p<0.01, * p<0.05

Coefficient Stability	CMS Overall Rating
Adjusted Beta	0.0968
Controlled Coefficient	0.0941
Controlled R Square	0.118
Max R Square	0.153

Discussion

Discussion of Findings

Prior studies mainly investigate the effects of online review factors on healthcare providers' medical performance or the correlations between them. In this essay, we argue that it is important to explore the impact of a hospital's medical performance on its online reputation, given that people first visit hospitals before they experience hospitals' medical quality and evaluate them online based on their perceptions. We adopt the coefficient stability method to handle possible omitted variable problems for the medical performance factors. The results from both methods are consistent. The significant results for the effects of readmission, mortality, safety of care, and time in ED support our central thesis that a hospital's medical performance significantly influences its online reputation.

Although patients and their families are unlikely to have the medical knowledge to evaluate the medical quality of treatment completely, it is still possible for them to gain some quality signals through their experience with the hospital. For example, families can observe whether their near and dear ones survive after the surgery in a hospital or not (Lu & Rui, 2018). Or, they know if the patient's condition deteriorates, they have to send her back to the hospital after a short time. Hence, it is reasonable that hospitals' medical performance factors, such as mortality and readmission, exert significant influence on their online ratings.

To the best of our knowledge, this study is the first attempt to investigate the effects of a representative set of medical performance factors on hospitals' online ratings. Offline survey scores are often treated as important indicators of hospitals' quality. However, this study focuses on objective quality measures. Thus, we exclude the offline survey score since it is a subjective evaluation by patients.

The robustness check further validates our findings. The overall hospital performance star ratings published by CMS has a lot of credibility, given that the federal government, the single largest health care payer in the U.S., issues them (Bilimoria & Barnard, 2021). Although prior studies have confirmed its influence on various medical outcomes for an extensive range of treatments, none of them has studied its effect on hospitals' online reputations. Our study fills the gap and contributes to the literature by unveiling the significant impact of the CMS overall quality star rating on online reputation.

In addition to the results for hospital medical performance, we find three influential factors from online review contents. First, we find that reviewer medical knowledge negatively affects online ratings. It means that if a reviewer demonstrates a higher medical knowledge level at the time she posts the review, which is reflected by using more complex and less frequent terms in her review, she is more likely to give a lower rating. Our study is the first attempt to explore the effect of reviewer medical knowledge on online ratings in the online healthcare review domain. The significant results suggest that a reviewer's medical knowledge conveyed in her review is a critical factor for review ratings.

Second, we find that if a reviewer mentions more information related to medical quality in her review, she will tend to give a higher online rating. On the other hand, if a review contains more service quality evaluations, the review gives a lower rating. We argue that it is because reviewers tend to have a higher expectation of service performance but a lower expectation of medical performance from a hospital because of their past experience. People tend to share experiences that exceed or fail their expectations. Specifically, we believe that reviewers are more likely to describe their medical experiences when those experiences surpass their expectations; as a result, the ratings will be higher. On the contrary, descriptions of service experience are more likely to be correlated with negative feelings. These results broaden our understanding of online healthcare reviews.

Managerial Implications

Our findings also provide significant practical implications to multiple participants of online healthcare reviews. First, our results for the four hospital medical performance factors provide hospitals with detailed information that will help them better understand their online reputation. For example, the results indicate that every additional 100-minute wait in the ED will lead to a drop of online rating by 0.37. Therefore, hospitals with long waiting times should pay attention to increasing their emergency department efficiency and reducing the average waiting time, so as to improve their online reputation.

The results of our study also suggest a significant positive relationship between the CMS overall quality star rating and review online rating. The CMS overall quality star rating can be treated as a reliable indicator since it can effectively reflect actual medical outcomes and a hospital's online reputation. Therefore, patients can use this composite performance measure as a benchmark for their decision-making processes. Moreover, hospitals should try to improve their performance with respect to the CMS overall quality star rating since it affects their online reputation, which has been shown to affect patient flows (Segal et al., 2012; Luca & Vats, 2013).

Prior studies find that product evaluations from experts are more persuasive (Petty et al., 1981) and reliable (Senecal & Nantel, 2004), thereby having more powerful impact (Vermeulen & Seegers, 2009). Given the importance of reviewers with medical knowledge, it is critical for hospitals to distinguish reviews that exhibit high medical knowledge from other reviews. By doing so, hospitals can mitigate the negative impact of the reviews written by knowledgeable reviewers. For example, most online review platforms allow businesses to respond to their online reviews. Tailoring manager responses to negative reviews has been an effective complaint management strategy, which positively influences subsequent opinions (Wang & Chaudhry, 2018). Hospitals can identify and respond to the reviews written by knowledgeable reviewers but with low ratings. If the tactic is applied appropriately, sincerely dealing with rational critics has the potential to enhance a hospital's image. Furthermore, since these knowledgeable reviewers are more likely to become opinion leaders and easily get dissatisfied, physicians and nurses should communicate with them promptly and try to assuage their negative feelings in the hospital, rather than let them express those feelings online.

The results for experiential quality evaluations are also helpful for hospitals to understand their patients. Low perceived service quality generates unfavorable behavioral intentions, which harm customer loyalty and cause financial losses. However, unfavorable feelings can be mitigated if the problems are resolved (Zeithaml et al., 1996). Objective performance is relatively stable and less likely to cause negative feelings. Therefore, it is crucial that hospitals handle reviews that express negative emotions with respect to service-related quality in a timely manner to reduce the likelihood of any negative consequences caused by dissatisfaction. For example, if a reviewer complains about the bad attitude of a physician at a hospital, the hospital could respond to the review and schedule a follow-up appointment. Meanwhile, hospitals should keep monitoring and tracking service-related issues expressed in online reviews and try to address them in the best possible way. Improving their online reputation through service quality initiatives is much more feasible in the short run than generating better scores on the objective performance measures.

Moreover, different impacts of experiential quality evaluations on online reputation are caused by different expectations. Therefore, reforming customers' expectations is a practical approach for hospitals to improve their online reputation. Golder et al. (2012) suggest that companies should seek to move expectations closer to the offerings' perceived attributes. Kopalle et al. (2017) find that it is not effective for a low-quality firm to emphasize quality in its advertising. In general, hospitals should reshape patient expectations so that their expectations are more aligned toward their medical performance, rather than their service performance. They could do so by emphasizing their competitive medical qualities in their advertisements.

Conclusions, Limitations, and Future Directions

Online healthcare reviews are becoming increasingly popular as reference sources for people making medical decisions since they are widely available and easy to understand. Given the significant information asymmetry and severe consequences of wrong decisions in the healthcare domain, it is important to understand what factors affect online healthcare reviews, especially the impact of hospital medical performance factors. In this study, we provide a comprehensive understanding of online healthcare reviews by investigating the effects of hospital medical performance from separate perspectives, as well as from an overall perspective, and extracting innovative and influential variables from review contents. The results show that most of the hospital medical performance factors significantly influence a hospital's online rating. We also find that the CMS overall quality star rating has a significant and positive impact on the hospital's online rating.

We employ machine learning techniques to extract helpful factors from review content. We find that reviewer medical knowledge has a significant and negative impact on online rating. In addition, the proportions of medical quality evaluations in a review positively affect online ratings, while service quality evaluations have a negative effect.

The measures of the experiential quality evaluations in this essay are extracted by applying an unsupervised learning method and classifying them based on subjective interpretations. The results support the necessity to investigate the effects of different experiential quality evaluations on online reputation. Future studies could develop specific customized medical vocabularies for different quality evaluation groups, such as service quality and medical quality. With those specialized vocabularies, researchers will be better equipped to come up with more accurate and effective topic classifications.

Furthermore, this study only investigates the effects of experiential quality evaluations' quantities on online ratings. However, we neglect the influences of different evaluations' sentiments. Future studies could apply text mining techniques to mine sentiments for various topics from review contents and study their correlations with overall numeric ratings.

This study explores the effects of objective hospital performance factors on their online reputation from a general perspective without considering different types of specialties or diseases. Future studies could further examine the effects of the proposed performance factors on online reputation in specific domains, such as chronic diseases.

References

1. Allard, T., Dunn, L. H., & White, K. (2020). Negative reviews, positive impact: Consumer empathetic responding to unfair word of mouth. *Journal of Marketing*, *84*(4), 86-108.
2. Allison, J. J., Kiefe, C. I., Weissman, N. W., Person, S. D., Rousculp, M., Canto, J. G., ... & Centor, R. M. (2000). Relationship of hospital teaching status with quality of care and mortality for Medicare patients with acute MI. *JAMA*, *284*(10), 1256-1262.
3. Andersson, J., Nordgren, L., Cheng, I., Nilsson, U., & Kurland, L. (2020). Long emergency department length of stay: A concept analysis. *International Emergency Nursing*, *53*, 100930.
4. Banerjee, S., Bhattacharyya, S., & Bose, I. (2017). Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business. *Decision Support Systems*, *96*, 17-26.
5. Bardach, N. S. (2018). Letter to the Editor in response to "Online physician ratings fail to predict actual performance on measures of quality, value, and peer review". *Journal of the American Medical Informatics Association*, *25*(6), 744-745.
6. Bardach, N. S., Asteria-Peñaloza, R., Boscardin, W. J., & Dudley, R. A. (2013). The relationship between commercial website ratings and traditional hospital performance measures in the USA. *BMJ Quality & Safety*, *22*(3), 194-202.
7. Bergman, S., Deban, M., Martelli, V., Monette, M., Sourial, N., Hamadani, F., ... & Fraser, S. (2014). Association between quality of care and complications after abdominal surgery. *Surgery*, *156*(3), 632-639.
8. Bettman, J. R., & Park, C. W. (1980). Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *Journal of Consumer Research*, *7*(3), 234-248.
9. Bilimoria, K. Y., & Barnard, C. (2021). An evolving Hospital quality StAR rating system from CMS: aligning the stars. *JAMA*, *325*(21), 2151-2152.
10. Bilimoria, K. Y., & Barnard, C. (2016). The new CMS hospital quality star ratings: the stars are not aligned. *JAMA*, *316*(17), 1761-1762.
11. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993-1022.
12. Bolton, R. N., & Drew, J. H. (1991). A longitudinal analysis of the impact of service changes on customer attitudes. *Journal of Marketing*, *55*(1), 1-9.
13. Bolton, R. N., & Drew, J. H. (1991). A multistage model of customers' assessments of service quality and value. *Journal of Consumer Research*, *17*(4), 375-384.
14. Bonis, P. A., Pickens, G. T., Rind, D. M., & Foster, D. A. (2008). Association of a clinical knowledge support system with improved patient safety, reduced complications and shorter length of stay among Medicare beneficiaries in acute care hospitals in the United States. *International Journal of Medical Informatics*, *77*(11), 745-753.
15. Boulding, W., Glickman, S. W., Manary, M. P., Schulman, K. A., & Staelin, R. (2011). Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days. *The American Journal of Managed Care*, *17*(1), 41-48.

16. Boulding, W., Kalra, A., & Staelin, R. (1999). The quality double whammy. *Marketing Science*, *18*(4), 463-484.
17. Boulding, W., Kalra, A., Staelin, R., & Zeithaml, V. A. (1993). A dynamic process model of service quality: from expectations to behavioral intentions. *Journal of Marketing Research*, *30*(1), 7-27.
18. Bowling, A., Rowe, G., Lambert, N., Waddington, M., Mahtani, K. R., Kenten, C., ... & Francis, S. A. (2012). The measurement of patients' expectations for health care: a review and psychometric testing of a measure of patients' expectations. *Health Technology Assessment*, *16*(30).
19. Brandão, C., Rego, G., Duarte, I., & Nunes, R. (2013). Social responsibility: a new paradigm of hospital governance?. *Health Care Analysis*, *21*(4), 390-402.
20. Brook, R. H., McGlynn, E. A., Shekelle, P. G., Marshall, M., Leatherman, S., & Adams, J. L. (2002). Report cards for health care: is anyone checking them. *RAND Corportation*.
21. Burke, L. G., Frakt, A. B., Khullar, D., Orav, E. J., & Jha, A. K. (2017). Association between teaching status and mortality in US hospitals. *JAMA*, *317*(20), 2105-2113.
22. Burkle, C. M., & Keegan, M. T. (2015). Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians. *BMC Health Services Research*, *15*(1), 1-7.
23. CDP (2019), All-cause admissions and readmissions. Spring 2018 cycle: CDP report. Technical Report, January 2019.
24. Chen, Y., & Lee, S. (2021). User-Generated Physician Ratings and Their Effects on Patients' Physician Choices: Evidence from Yelp. *Available at SSRN 3796740*.
25. Cheng, Y. H., & Ho, H. Y. (2015). Social influence's impact on reader perceptions of online reviews. *Journal of Business Research*, *68*(4), 883-887.
26. Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345-354.
27. Cordell, V. V. (1997). Consumer knowledge measures as predictors in product evaluation. *Psychology & Marketing*, *14*(3), 241-260.
28. Curry, D. J., & Riesz, P. C. (1988). Prices and price/quality relationships: a longitudinal analysis. *Journal of Marketing*, *52*(1), 36-51.
29. Daskivich, T. J., Houman, J., Fuller, G., Black, J. T., Kim, H. L., & Spiegel, B. (2018). Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. *Journal of the American Medical Informatics Association*, *25*(4), 401-407.
30. Doyle, C., Lennox, L., & Bell, D. (2013). A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open*, *3*(1), e001570.
31. Fontana, M. A., Lyman, S., Islam, W., & MacLean, C. H. (2019). When stars do not align: overall hospital quality star ratings and the volume-outcome association. *JBJS Open Access*, *4*(1).
32. Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*(3), 291-313.

33. Gao, G. G., McCullough, J. S., Agarwal, R., & Jha, A. K. (2012). A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. *Journal of Medical Internet Research*, *14*(1), e2003.
34. Gao, G., Greenwood, B. N., Agarwal, R., & McCullough, J. S. (2015). Vocal minority and silent majority. *MIS Quarterly*, *39*(3), 565-590.
35. Geraci, J. M., Ashton, C. M., Kuykendall, D. H., Johnson, M. L., Souchek, J., Del Junco, D., & Wray, N. P. (1999). The association of quality of care and occurrence of in-hospital, treatment-related complications. *Medical Care*, 140-148.
36. Ghose, A., & Ipeirotis, P. G. (2010). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498-1512.
37. Glickman, S. W., Boulding, W., Manary, M., Staelin, R., Roe, M. T., Wolosin, R. J., ... & Schulman, K. A. (2010). Patient satisfaction and its relationship with clinical quality and inpatient mortality in acute myocardial infarction. *Circulation: Cardiovascular Quality and Outcomes*, *3*(2), 188-195.
38. Glover, M., Khalilzadeh, O., Choy, G., Prabhakar, A. M., Pandharipande, P. V., & Gazelle, G. S. (2015). Hospital evaluations by social media: a comparative analysis of Facebook ratings among performance outliers. *Journal of General Internal Medicine*, *30*(10), 1440-1446.
39. Golder, P. N., Mitra, D., & Moorman, C. (2012). What is quality? An integrative framework of processes and states. *Journal of Marketing*, *76*(4), 1-23.
40. Gray, B. M., Vandergrift, J. L., Gao, G. G., McCullough, J. S., & Lipner, R. S. (2015). Website ratings of physicians and their quality of care. *JAMA Internal Medicine*, *175*(2), 291-293.
41. Greaves, F., Laverty, A. A., Cano, D. R., Moilanen, K., Pulman, S., Darzi, A., & Millett, C. (2014). Tweets about hospital quality: a mixed methods study. *BMJ Quality & Safety*, *23*(10), 838-846.
42. Greaves, F., Pape, U. J., Lee, H., Smith, D. M., Darzi, A., Majeed, A., & Millett, C. (2012). Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. *Journal of Medical Internet Research*, *14*(5), e2280.
43. Guo, B., & Zhou, S. (2016). Understanding the impact of prior reviews on subsequent reviews: The role of rating volume, variance and reviewer characteristics. *Electronic Commerce Research and Applications*, *20*, 147-158.
44. Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). Factor analysis. Multivariate data analysis. *NJ Prentice-Hall*, *3*, 98-99.
45. Han, M. (2021). Examining the effect of reviewer expertise and personality on reviewer satisfaction: an empirical study of TripAdvisor. *Computers in Human Behavior*, *114*, 106567.
46. Hanauer, D. A., Zheng, K., Singer, D. C., Gebremariam, A., & Davis, M. M. (2014). Public awareness, perception, and use of online physician rating sites. *JAMA*, *311*(7), 734-735.
47. Hedges, L., & Couey, C. (2019). How patients use online reviews. *Software Advice*.

48. Horwitz, L., Partovian, C., Lin, Z., Herrin, J., Grady, J., Conover, M., ... & Krumholz, H. M. (2011). Hospital-wide (all-condition) 30-day risk-standardized readmission measure. *New Haven, CT*.
49. Johnson, E. J., & Russo, J. E. (1984). Product familiarity and learning new information. *Journal of Consumer Research*, *11*(1), 542-550.
50. Johnson, M. D., Anderson, E. W., & Fornell, C. (1995). Rational and adaptive performance expectations in a customer satisfaction framework. *Journal of Consumer Research*, *21*(4), 695-707.
51. Kamakura, W. A., Mittal, V., De Rosa, F., & Mazzon, J. A. (2002). Assessing the service-profit chain. *Marketing Science*, *21*(3), 294-317.
52. Kassirer, J. P. (1999). Hospitals, heal yourselves. *New England Journal of Medicine*, *340*(4), 309-310.
53. Kennedy, P. (2008). *A guide to econometrics*. John Wiley & Sons.
54. Khan, N. A., Quan, H., Bugar, J. M., Lemaire, J. B., Brant, R., & Ghali, W. A. (2006). Association of postoperative complications with hospital costs and length of stay in a tertiary care center. *Journal of General Internal Medicine*, *21*(2), 177-180.
55. Kim, E. E. K., Mattila, A. S., & Baloglu, S. (2011). Effects of gender and expertise on consumers' motivation to read online hotel reviews. *Cornell Hospitality Quarterly*, *52*(4), 399-406.
56. Koh, C. Y., Inaba, C. S., Sujatha-Bhaskar, S., & Nguyen, N. T. (2017). Association of Centers for Medicare & Medicaid Services overall hospital quality star rating with outcomes in advanced laparoscopic abdominal surgery. *JAMA Surgery*, *152*(12), 1113-1117.
57. Kolstad, J. T. (2013). Information and quality when motivation is intrinsic: Evidence from surgeon report cards. *American Economic Review*, *103*(7), 2875-2910.
58. Kopalle, P. K., Fisher, R. J., Sud, B. L., & Antia, K. D. (2017). The effects of advertised quality emphasis and objective quality on sales. *Journal of Marketing*, *81*(2), 114-126.
59. Kordzadeh, N. (2019). Investigating bias in the online physician reviews published on healthcare organizations' websites. *Decision Support Systems*, *118*, 70-82.
60. Ladhari, R., Pons, F., Bressolles, G., & Zins, M. (2011). Culture and personal values: How they influence perceived service quality. *Journal of Business Research*, *64*(9), 951-957.
61. Landon, S., & Smith, C. E. (2012). Quality expectations, reputation, and price. In *World Scientific Reference on Handbook of the Economics of Wine: Volume 2: Reputation, Regulation, and Market Organization* (pp. 3-31).
62. LaTour, S. A., & Peat, N. C. (1979). Conceptual and methodological issues in consumer satisfaction research. *ACR North American Advances*.
63. LaTour, S. A., & Peat, N. C. (1980). The role of situationally-produced expectations, others' experiences, and prior experience in determining consumer satisfaction. *ACR North American Advances*.
64. Lawrence, B., & Perrigot, R. (2015). Influence of organizational form and customer type on online customer satisfaction ratings. *Journal of Small Business Management*, *53*, 58-74.
65. Lee, J. Y., Gowen III, C. R., & McFadden, K. L. (2018). An empirical study of US hospital quality: Readmission rates, organizational culture, patient satisfaction, and Facebook ratings. *Quality Management Journal*, *25*(4), 158-170.

66. Lichtenstein, D. R., & Burton, S. (1989). The relationship between perceived and objective price-quality. *Journal of Marketing Research*, *26*(4), 429-443.
67. Lu, S. F., & Rui, H. (2018). Can we trust online physician ratings? Evidence from cardiac surgeons in Florida. *Management Science*, *64*(6), 2557-2573.
68. Luca, M., & Vats, S. (2013). Digitizing doctor demand: The impact of online reviews on doctor choice. *Cambridge, MA: Harvard Business School*.
69. Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing*, *77*(1), 87-103.
70. Maheswaran, D., Sternthal, B., & Gürhan, Z. (1996). Acquisition and impact of consumer expertise. *Journal of Consumer Psychology*, *5*(2), 115-133.
71. McCallum, A. K. (2002). Mallet: A machine learning for language toolkit. *http://mallet.cs.umass.edu*.
72. Mehta, R., Paredes, A. Z., Tsilimigras, D. I., Farooq, A., Sahara, K., Merath, K., ... & Pawlik, T. M. (2020). CMS hospital compare system of star ratings and surgical outcomes among patients undergoing surgery for cancer: do the ratings matter?. *Annals of Surgical Oncology*, *27*(9), 3138-3146.
73. Mitra, D., & Golder, P. N. (2006). How does objective quality affect perceived quality? Short-term effects, long-term effects, and asymmetries. *Marketing Science*, *25*(3), 230-247.
74. Montes, F. R., Vásquez, S. M., Camargo-Rojas, C. M., Rueda, M. V., Góez-Mogollón, L., Alvarado, P. A., ... & Villar, J. C. (2019). Association between emergency department length of stay and adverse perioperative outcomes in emergency surgery: a cohort study in two Colombian University hospitals. *BMC Emergency Medicine*, *19*(1), 1-10.
75. Luo, L., Duan, S., Shang, S., & Pan, Y. (2021). What makes a helpful online review? Empirical evidence on the effects of review and reviewer characteristics. *Online Information Review*.
76. Mullins, P. M., & Pines, J. M. (2014). National ED crowding and hospital quality: results from the 2013 Hospital Compare data. *The American Journal of Emergency Medicine*, *32*(6), 634-639.
77. Ngo-Ye, T. L., & Sinha, A. P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, *61*, 47-58.
78. Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, *17*(4), 460-469.
79. Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187-204.
80. Papageorge, M. V., Resio, B. J., Monsalve, A. F., Canavan, M., Pathak, R., Mase, V. J., ... & Boffa, D. J. (2020). Navigating by stars: using CMS Star Ratings to choose hospitals for complex cancer surgery. *JNCI Cancer Spectrum*, *4*(5), pkaa059.
81. Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, *49*(4), 41-50.
82. Park, C. W., & Lessig, V. P. (1981). Familiarity and its impact on consumer decision biases and heuristics. *Journal of Consumer Research*, *8*(2), 223-230.

83. Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, *41*(5), 847.
84. Prabhu, J., & Tellis, G. J. (2000). Do consumers ever learn? Analysis of segment behavior in experimental markets. *Journal of Behavioral Decision Making*, *13*(1), 19-34.
85. Ranard, B. L., Werner, R. M., Antanavicius, T., Schwartz, H. A., Smith, R. J., Meisel, Z. F., ... & Merchant, R. M. (2016). Yelp reviews of hospital care can supplement and inform traditional surveys of the patient experience of care. *Health Affairs*, *35*(4), 697-705.
86. Rastegar-Mojarad, M., Ye, Z., Wall, D., Murali, N., & Lin, S. (2015). Collecting and analyzing patient experiences of health care from social media. *JMIR Research Protocols*, *4*(3), e3433.
87. Rehurek, R., & Sojka, P. (2011). Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, *3*(2), 2.
88. Reinartz, W., & Kumar, V. I. S. W. A. N. A. T. H. A. N. (2002). The mismanagement of customer loyalty. *Harvard Business Review*, *80*(7), 86-94.
89. Riesz, P. C. (1978). Price versus quality in the marketplace, 1961-1975. *Journal of Retailing*, *54*(4), 15-28.
90. Sahoo, N., Dellarocas, C., & Srinivasan, S. (2018). The impact of online product reviews on product returns. *Information Systems Research*, *29*(3), 723-738.
91. Saifee, D. H., Bardhan, I. R., Lahiri, A., & Zheng, Z. (2019). Adherence to clinical guidelines, electronic health record use, and online reviews. *Journal of Management Information Systems*, *36*(4), 1071-1104.
92. Saifee, D. H., Zheng, Z., Bardhan, I. R., & Lahiri, A. (2020). Are online reviews of physicians reliable indicators of clinical outcomes? A focus on chronic disease management. *Information Systems Research*, *31*(4), 1282-1300.
93. Salinas, S. R. (2017). Examining the relationship between perceived quality of care and actual quality of care as measured by 30-day readmission rates. *Quality Management in Health Care*, *26*(1), 29-32.
94. Segal, J., Sacopulos, M., Sheets, V., Thurston, I., Brooks, K., & Puccia, R. (2012). Online doctor reviews: do they track surgeon volume, a proxy for quality of care?. *Journal of Medical Internet Research*, *14*(2), e2005.
95. Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, *80*(2), 159-169.
96. Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, *20*(4), 518-529.
97. Vanderpool, D. (2017). What can I do about a negative online review?. *Innovations in Clinical Neuroscience*, *14*(5-6), 31.
98. Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, *30*(1), 123-127.
99. Wang, Y., & Chaudhry, A. (2018). When and how managers' responses to online reviews affect subsequent reviews. *Journal of Marketing Research*, *55*(2), 163-177.

100. Werner, R. M., Norton, E. C., Konetzka, R. T., & Polsky, D. (2012). Do consumers respond to publicly reported quality information? Evidence from nursing homes. *Journal of Health Economics*, *31*(1), 50-61.
101. Woodruff, R. B., Cadotte, E. R., & Jenkins, R. L. (1983). Modeling consumer satisfaction processes using experience-based norms. *Journal of Marketing Research*, *20*(3), 296-304.
102. Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
103. Xu, Y., Armony, M., & Ghose, A. (2021). The interplay between online reviews and physician demand: An empirical investigation. *Management Science*, *67*(12), 7344-7361.
104. Yaraghi, N., Wang, W., Gao, G. G., & Agarwal, R. (2018). How online quality ratings influence patients' choice of medical providers: Controlled experimental survey study. *Journal of Medical Internet Research*, *20*(3), e8986.
105. Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, *38*(2), 539-560.
106. Yin, D., Bond, S. D., & Zhang, H. (2017). Keep your cool or let it out: Nonlinear effects of expressed arousal on perceptions of consumer reviews. *Journal of Marketing Research*, *54*(3), 447-463.
107. Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *Journal of Marketing*, *52*(3), 2-22.
108. Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, *60*(2), 31-46.
109. Zhang, J., & Liu, R. R. (2019). The more the better? Exploring the effects of reviewer social networks on online reviews. *Journal of Marketing Management*, *35*(17-18), 1667-1688.
110. Zhu, L., Yin, G., & He, W. (2014). Is this opinion leader's review useful? Peripheral cues for online review helpfulness. *Journal of Electronic Commerce Research*, *15*(4), 267.
111. Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, *74*(2), 133-148.
112. Zou, P., Yu, B., & Hao, Y. (2011). Does the valence of online consumer reviews matter for consumer decision making? The moderating role of consumer expertise. *Journal of Computers*, *6*(3), 484-488.

Appendix: Correlation Matrix

CMS Overall Rating	1.00	-0.32	-0.25	-0.04	-0.15	0.01	0.00	0.00	0.01	0.02	0.04	0.00	0.01	-0.01	0.02	0.09
Readmission	-0.32	1.00	-0.35	0.17	0.15	0.00	0.01	0.00	-0.01	-0.02	-0.02	0.00	0.00	-0.01	-0.01	0.02
Mortality	-0.25	-0.35	1.00	-0.23	-0.26	-0.01	0.01	0.00	0.00	0.01	0.02	0.00	-0.01	0.01	0.01	-0.17
Safety of Care	-0.04	0.17	-0.23	1.00	0.12	0.00	0.00	0.00	-0.01	-0.02	-0.02	0.00	0.02	-0.02	-0.01	0.04
Time in ED	-0.15	0.15	-0.26	0.12	1.00	0.00	-0.02	0.00	0.00	0.00	-0.02	0.01	0.02	-0.01	0.00	0.32
Reviewer Medial Knowledge	0.01	0.00	-0.01	0.00	0.00	1.00	0.12	-0.13	-0.01	-0.02	-0.03	-0.01	0.00	0.01	-0.01	-0.01
Medical Quality Evaluations	0.00	0.01	0.01	0.00	-0.02	0.12	1.00	-0.51	-0.03	-0.04	-0.05	-0.03	0.03	0.03	-0.03	-0.04
Service Quality Evaluations	0.00	0.00	0.00	0.00	0.00	-0.13	-0.51	1.00	-0.03	-0.04	-0.04	-0.03	0.00	-0.04	0.00	0.01
Reviewer Number of Friends	0.01	-0.01	0.00	-0.01	0.00	-0.01	-0.03	-0.03	1.00	0.47	0.42	0.62	0.04	0.01	-0.01	0.02
Reviewer Number of Reviews	0.02	-0.02	0.01	-0.02	0.00	-0.02	-0.04	-0.04	0.47	1.00	0.52	0.67	0.04	0.01	0.02	0.03
Reviewer Number of Badges	0.04	-0.02	0.02	-0.02	-0.02	-0.03	-0.05	-0.04	0.42	0.52	1.00	0.34	0.09	0.01	0.03	0.05
Reviewer Number of Followers	0.00	0.00	0.00	0.00	0.01	-0.01	-0.03	-0.03	0.62	0.67	0.34	1.00	0.04	0.00	0.01	0.02
Review Length	0.01	0.00	-0.01	0.02	0.02	0.00	0.03	0.00	0.04	0.04	0.09	0.04	1.00	-0.16	0.23	0.03
Review Subjectivity	-0.01	-0.01	0.01	-0.02	-0.01	0.01	0.03	-0.04	0.01	0.01	0.01	0.00	-0.16	1.00	-0.08	-0.02
Review Fog Index	0.02	-0.01	0.01	-0.01	0.00	-0.01	-0.03	0.00	-0.01	0.02	0.03	0.01	0.23	-0.08	1.00	0.02
Hospital Number of Beds	0.09	0.02	-0.17	0.04	0.32	-0.01	-0.04	0.01	0.02	0.03	0.05	0.02	0.03	-0.02	0.02	1.00

Essay III:

The Determinants of Hospitals' Online Reputations: Mining Patients' Evaluations from Online Healthcare Reviews Using an Aspect-Based Sentiment Analysis

Abstract

Due to the complex nature of healthcare services, it is difficult for patients who lack medical knowledge to accurately evaluate the quality of service they receive from healthcare providers. Because of this information asymmetry, patients largely rely on various information sources, such as word of mouth, to make medical decisions. In recent years, as online reviews are getting increasingly popular, online healthcare reviews have become one of the most common reference sources for patients. Most online healthcare review studies are at the document level, investigating the reviews' numeric ratings or overall sentiment polarity. However, we argue that it is necessary to mine deeper into online healthcare reviews to an aspect level to help providers understand their online reputations and facilitate accurate referral of patients. This essay proposes a hybrid framework that combines aspect-based sentiment analysis (ABSA) and regression models to mine patients' online evaluations of a hospital from different aspects and study the effects of aspect categories' sentiments on online ratings. The results show that including the aspect categories' polarities dramatically increases model fit. The standardized coefficients reveal that "Staff," "Nurse," and "Doctor" are the three most influential aspects of hospitals' online review ratings. The findings from this essay underscore the need to adopt ABSA in the online healthcare review domain. They also have practical implications for patients, healthcare providers, and online review platforms.

Keywords: online healthcare reviews, aspect-based sentiment analysis, text mining, online ratings, topic modeling

Introduction

Due to the rapid development of electronic commerce in recent years, the impact of online customer reviews on various key aspects, such as online reputation, customer decision, and sales, of a business have become increasingly crucial (Chevalier & Mayzlin, 2006; Ludwig et al., 2013; Sahoo et al., 2018; Zhu & Zhang, 2010). Online customer reviews, one of the most popular channels for expressing product evaluations, contain customers' subjective opinions on products or services. Therefore, they become reliable references for potential customers for making their purchase decisions.

The reference role of online reviews is even more vital in the healthcare domain due to its information asymmetry characteristic, where patients typically lack the specialized knowledge required to evaluate the quality of the service (Arrow, 1978). A recent survey finds that 72% of patients first refer to online reviews when looking for a new doctor (Hedges & Couey, 2019). A typical online review has two parts – the numeric rating and textual content, where the numeric rating represents a reviewer's overall evaluation of the product or service, and the content has detailed information. Previous online healthcare review studies mostly investigate numeric ratings or overall sentiment polarities from review contents (e.g., Luca & Vats; 2013, Saifee et al., 2019). However, only looking at numeric ratings or overall sentiments ignores a sufficient level of detail or the specific aspects of the product or service that customers have evaluated. (D'Aniello et al., 2022). Therefore, when various aspects are involved, capturing the sentiments associated with different aspects is necessary, especially for complicated products or services, such as healthcare services. To mine fine-grained opinions over coarse-grained opinions, an innovative technique called aspect-based sentiment analysis (ABSA) has been developed (Do et al., 2019; Zhao et al., 2016).

ABSA identifies a reviewer's sentiments on different aspects embedded in the review content (Eirinaki et al., 2012). An aspect usually represents an item of a specific topic in a domain, such as doctor, nurse, staff, treatment, and diagnosis, in the healthcare domain. A typical ABSA consists of three main tasks: identifying aspects from a document, extracting descriptive expression for aspects, and calculating sentiments for aspects (Pontiki et al., 2016). ABSA techniques have been adopted widely in recent years since more detailed information, rather than overall polarities, is required in many situations (Moghaddam & Ester, 2013; Mukherjee & Liu, 2012). The implementations of ABSA techniques can be found in many domains, such as movies (Thet et al., 2010), digital products (Hu & Liu, 2004), and restaurants (Ganu et al., 2009; Brody & Elhadad, 2010).

Due to the complex nature of healthcare services (Engelseth et al., 2021; Rouse & Serban, 2014), online healthcare reviews should contain evaluations of healthcare providers from various aspects, emphasizing the need for adopting ABSA techniques in this domain. However, to our knowledge, no study has specifically investigated the implementation of ABSA techniques in online healthcare reviews.

This essay proposes a framework for implementing ABSA for online healthcare reviews to extract sentiments of the underlying aspect categories and investigate their effects on online ratings. The proposed ABSA combines vocabulary-based, topic model-based, and syntactic relation-based ABSA techniques. Meanwhile, it overcomes the major limitations of the three types of ABSA techniques. The framework consists of five major tasks: aspect categories identification, development of dictionaries, data preparation, sentiments analysis, and data analysis. Specifically, we identify a set of aspect categories from existing surveys and studies. Next, we supplement additional aspect categories using the results from topic modeling. We extract 18 aspect categories, such as “Doctor,” “Nurse,” “Staff,” and “Diagnosis,” which form a holistic view of patients' evaluations. In the next step, we develop dictionaries for the identified aspect categories. The pre-defined dictionaries contain core and relevant terms of the corresponding categories. More importantly, we further enrich the dictionaries with the matched Consumer Health Vocabulary (CHV) terms in the reviews so that the pre-defined dictionaries are more relevant to the target online reviews.

After defining the dictionaries for the 18 aspect categories, we conduct several typical data cleansing steps, such as stop word removal, tokenization, and lemmatization. We further perform an additional step of replacing the dictionaries' phrases with their n-grams in reviews. This step increases the accuracy of the Part-of-Speech (POS) tags and token dependencies. Next, we perform sentiment analysis at the sentence level by matching the keywords and identifying the corresponding opinion terms. The outputs from the ABSA are the average aspect category sentiment scores. Finally, we include the outputs into regression models to examine the determinants of hospitals' online ratings.

To illustrate the need for adopting ABSA in the online healthcare review domain and the usefulness of our proposed framework, we implement the framework on a dataset collected from Yelp.com. The final dataset contains 29,432 reviews for 686 hospitals from 49 states across the US from 2016 to 2020. We first extract the average sentiments from the online reviews. Next, we integrate the sentiment scores in regression models to investigate their effects on numeric online ratings. The empirical results indicate that most of the aspect categories' sentiments positively affect online ratings, and including the sentiment scores significantly improves model fit. Furthermore, the standardized coefficients of the aspect categories can reveal the categories' relative importance to online ratings. Based on the results, the three most important categories for hospitals' online ratings are "Staff," "Nurse," and "Doctor."

This essay contributes to both theory and practice. First, despite the importance of online healthcare reviews and the complexity of healthcare services, no study has specifically investigated how to implement ABSA approaches in online healthcare reviews. To the best of our knowledge, this essay is the first to propose a comprehensive framework that combines ABSA and econometric models to identify online ratings' determinants for online healthcare reviews. With respect to practice, the results of this essay can help hospitals better understand their online evaluations, thereby helping them improve their reputations. The results can also facilitate patients accurately referring to online reviews to make medical decisions. Furthermore, we recommend that online review platforms display sentiment scores for different aspects extracted from the contents to provide a comprehensive image of a hospital's online reputation.

The rest of the essay is arranged as follows. We first review the relevant literature related to online healthcare reviews and ABSA. Next, we introduce our proposed framework in detail. We then describe our data and variables, followed by the empirical analysis and results. Finally, we discuss the contributions and implications of our findings, and conclude the essay by pointing out the limitations and future directions.

Literature Review

This section first reviews existing online healthcare reviews studies that investigate online review contents using text mining techniques, such as topic modeling. Next, we review the online review studies that apply the aspect-based sentiment analysis approach.

Online Healthcare Reviews

With the importance of online reviews being increasingly realized by healthcare providers and patients, researchers have started to investigate online reviews in the healthcare domain from different angles (e.g., Gao et al., 2012; Kordzadeh, 2019).

A stream of research focuses on examining correlations between healthcare providers' online ratings and various performance measures, such as the HCAHPS survey (Ranard et al., 2016), administrative survey (Daskivich et al., 2018), consumer checkbook (Gao et al., 2015), physician certifications and school rankings (Gao et al., 2012), mortality rates (Lu & Rui, 2018), and process-of-care composite (Gray et al., 2015). For example, Saifee et al. (2020) study the correlation between physicians' online review ratings and patients' clinical outcomes, such as readmission risk and emergency room visits, but they do not find any significant relationships. On the other hand, Chen and Lee (2021) find that online ratings are positively and significantly associated with clinical measures, including physicians' adherence to clinical guidelines and patients' risk-adjusted health outcomes.

Another research stream investigates the relationship between online reviews and providers' demands or patient choices (e.g., Segal et al., 2012; Xu et al., 2021). Luca and Vats (2013) study how online review ratings affect patient demand. Using regression discontinuity, they find that every half a star improvement in ratings leads to a 10% increase in the average probability that a doctor will fill an appointment.

Despite the fact that the popularity of online healthcare reviews has been growing rapidly, the studies that mine deeper into review content using text mining techniques are still lacking. Using Natural Language Processing (NLP) tools and sentiment analysis techniques, Rastegar-Mojarad et al. (2015) develop a corpus of patient experience (COPE). COPE contains

79,173 sentences from 6,914 patient reviews of 985 healthcare providers for near 30 universities in the US. It is developed to be used to extract knowledge of patient experience from their evaluations. Saifee et al. (2019) extract latent topics from online healthcare review contents using the topic modeling approach and study their effects on online review ratings. They identify four latent topics: surgery, staff, physician, and overall care. They find that the proportions of surgery and overall care positively affect review ratings. However, more contribution from the staff topic hurts the ratings. Other than the numeric review ratings, Saifee et al. (2019) also investigate the overall textual sentiments; the results are consistent with the numeric rating models.

Although exploiting information from review content is essential, the topic modeling approach and general sentiment analysis still have limitations. First, the topic modeling approach only generates proportions of latent topics but not their corresponding sentiments. Moreover, using sentiment analysis to extract the overall polarity of a review does not differentiate reviewers' attitudes toward different aspects of a product or service. Therefore, it is necessary to have a method that mines reviewers' opinions for each specific aspect of a product or service. This kind of analysis method is known as aspect-based sentiment analysis (ABSA).

ABSA and Online Reviews

Hu and Liu (2004) first proposed the ABSA method and, since then, it has drawn a great deal of attention from researchers. ABSA is vital for extracting high-granular sentiments (Sarawgi & Pathak, 2017). Specifically, unlike traditional sentiment analysis, which treats a document as a unit of measure and generates an overall polarity or sentiment category (positive, negative, or neutral), the ABSA approach extracts multiple aspects from one document and calculates the corresponding sentiments for each of them. ABSA involves two primary tasks: aspect detection and sentiment analysis. ABSA has been applied to many domains, such as recommendation systems (Osman, 2019), domain ontology (Al-Aswadi et al., 2020), tourism (Alaei et al., 2019), online education (Kastrati et al., 2020), and transportation (Ali et al., 2019). Due to the richness of textual contents, the online customer review domain has become a popular area that applies ABSA methods (e.g., Fang & Tao, 2019; Loke & Reitter, 2021; Xue et al., 2017; Zhao et al., 2021).

A popular study stream follows the design science research paradigm (Gregor & Hevner, 2013) and proposes innovative ABSA methods as IT artifacts and evaluate them in certain cases to prove the superiority of their proposed methods. (e.g., Banjar, 2021; D'Aniello et al., 2022; Feng et al., 2022; Qiu et al., 2011). Chen and Xu (2017) propose an innovative ABSA method that combines product ontology and topic modeling. They illustrate the usefulness of their proposed framework on online reviews for single-lens reflex cameras. The proposed model identifies eight specific aspects and the results reveal that cost performance, image quality, and product integrity are the three most influential aspects of online review ratings. Similarly, Al-Ghuribi et al. (2020) propose an ABSA approach for large-scale user reviews and implement the proposed approach on Amazon and Yelp reviews in the domains of book, movie, and restaurant. Another paper (Cheng & Yang, 2022) focuses on the effect of online reviews on movie box office sales. They propose a framework that integrates ABSA and econometric models. Five aspects are defined from the proposed framework, including overall impression, screenplay, special effects, director, and principal characters. The empirical results indicate that regardless of sentiments, the five aspects positively impact movie revenues.

The adoption of ABSA methods in the online healthcare review domain is still in its early stage. Only a few studies introduce ABSA-relevant concepts in the online healthcare review domain. Xu et al. (2021) investigate how online reviews affect physician demand. They derive seven service-quality measures from review contents using the ABSA approach. The top seven most frequent features are bedside manner, accuracy of diagnosis, waiting time, service time, ease of insurance process, physician's knowledge, and office cleanliness. The empirical results show that bedside manner, accuracy of diagnosis, waiting time, and service time positively and significantly affect physicians' demands. To compare sentiment on Twitter about hospitals with established survey measures of patient experience, Greaves et al. (2014) manually identify six key themes from tweets, including quality, fundraising activities, health information, organizational or practical information about the hospital, promotional messages, and messages to patients receiving care. They find that 77% of tweets about care quality are positive. However, there is no association between Twitter sentiment and conventional quality metrics. Although these two studies involve ABSA to some extent, they are essentially different from this essay.

First, none of them studies the impacts of different aspects' sentiments on online review ratings. The numeric review ratings represent reviewers' overall evaluations of hospitals. Therefore, unveiling how different aspects' sentiments affect online review ratings helps hospitals understand their online reputations effectively. Second, we apply different approaches to extract aspects. Xu et al. (2021) identify aspects based on keywords' frequency. The limitations of the frequency-based approach are that it may select words that are not aspects and ignore aspects that are not frequently mentioned (Al-Ghuribi et al., 2020). Greaves et al. (2014) retrieve aspects by manual coding. This approach is not efficient when the number of documents is large. This essay proposes an ABSA framework that combines vocabulary-based, syntactic relation-based, and topic model-based techniques. It can overcome the major limitations of the previous ABSA approaches.

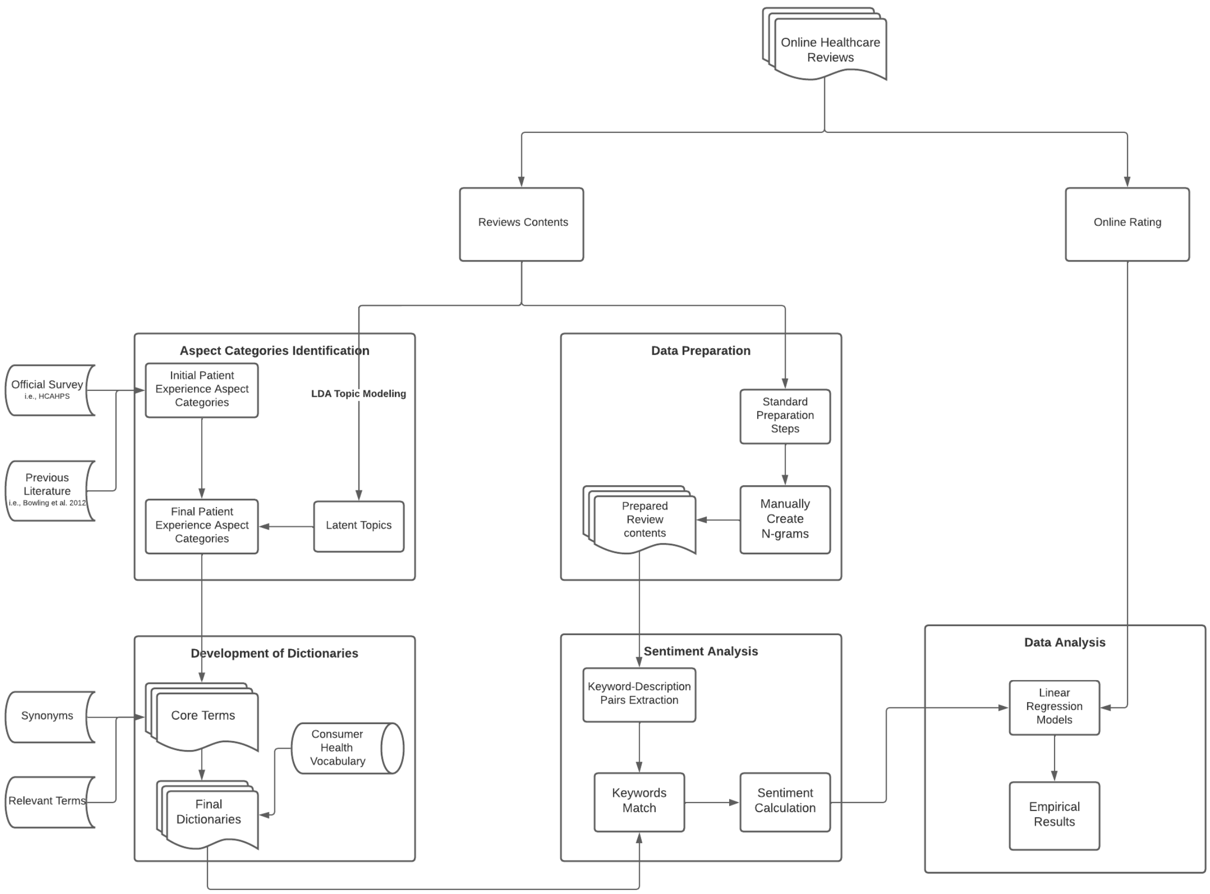
Aspect-Based Sentiment Analysis for Online Healthcare Reviews

This section introduces the framework for extracting aspect categories' sentiments embedded in online healthcare reviews and examining their effects on online ratings. The proposed framework includes five tasks: aspect categories identification, development of dictionaries, data preparation, sentiment analysis, and data analysis. All these works are implemented by Python. Figure 3.1 presents an overview of the framework.

As Figure 3.1 shows, we first inherit some patient experience classifications from the previous studies (i.e., Bowling et al., 2012) and official survey questions (i.e., HCAHPS Survey). Next, we employ the topic modeling method to discover possible supplementary aspect categories from online review contents. In sum, we identify 18 aspect categories. After determining the aspect categories, we develop a dictionary for each category. Each dictionary contains synonyms of the core terms and the vocabularies in the same context. We also enrich the dictionaries by mapping the matched Consumer Healthcare Vocabulary CHV terms into appropriate categories.

The data preparation step involves standard operations, including tokenization, stop word removal, lemmatization, bigram, and trigram creation. Besides, we introduce an additional step to replace the phrases having more than one word with n-grams. This step can increase the accuracy of the Part-of-Speech (POS) tags and the token dependencies. In the sentiment analysis step, we extract the keyword-description pairs using the POS tags and token dependencies and then match the keywords with terms in the dictionaries. We calculate the sentiment polarities for the corresponding descriptive terms of the matched keywords and take the average of the sentiment scores for all descriptive terms from the same aspect category. Finally, we integrate the aspect categories' sentiments in regression models with the online rating as the dependent variable. The analysis results can provide vital information for the determinants of healthcare providers' online ratings. All the steps are discussed in detail in the following subsections.

Figure 3. 1 The Framework of Applying ABSA for Online Healthcare Reviews



Aspect Categories Identification

Many prior studies use unsupervised learning models to extract aspects and build dictionaries. The unsupervised methods for ABSA can be classified into four categories (Hernández-Rubio et al., 2019): vocabulary-based (i.e., Siering et al., 2018), frequency-based (i.e., Dragoni et al., 2019), syntactic relation-based (Qiu et al., 2011), and topic model-based (Lin & He, 2009) methods. However, different types of ABSA approaches suffer from different limitations. For example, frequency-based approaches select aspects that solely rely on term occurrence. As a result, these approaches commonly have two major limitations: first, many extracted aspects are not relevant to the domains; second, infrequent nouns are ignored despite their importance to the domains (Rana & Cheah, 2017). Similarly, syntactic relation-based methods, which analyze the syntactic structure of sentences and relations among words to identify aspect's sentiment words, also suffer from these two limitations (Tubishat et al., 2021).

The two drawbacks mentioned above are especially significant in the healthcare domain due to the complex nature of healthcare services. They involve more components than common products or services, such as digital products, hotels, and restaurants. Therefore, evaluations of healthcare services should contain more aspects than other products or services, as evidenced from the number of aspects extracted from online reviews by previous studies. For example, Chen and Xu (2017) combine product ontology and topic modeling approach and extract eight aspects of a single-lens reflex camera. Chen and Yang (2022) investigate movies' online reviews and define five aspects. In comparison, we retrieve 18 aspects of online healthcare services. It is challenging to cover all aspects of healthcare services solely relying on term occurrences since some essential medical aspects, such as referral or recovery, may not be frequently mentioned in reviews. As a result, they cannot be identified by unsupervised learning models.

Furthermore, healthcare reviews contain many medical terminologies for aspects such as medicine, symptom, disease, medical test, and treatment. Due to the nature of user-generated reviews, reviewers may describe their experiences with different words without following a standard. Thus, reviewers may use different expressions to deliver similar intentions and meanings (Al-Ghuribi et al., 2020). Therefore, occurrences of individual medical terms in online healthcare reviews could be low, but they may belong to the same categories. For example, each of the 100 different medicine names occurs in reviews five times. According to the frequencies, none of them will be recognized. However, they represent 500 occurrences of the "medicine" aspect in total. It is also true for the medical test abbreviations. The two limitations of frequency-based and syntactic relation-based ABSA can be resolved by pre-defining the aspect categories and developing dictionaries.

In this essay, we identify 18 aspect categories. Fifteen are summarized from official surveys and previous studies, and three additional topics are supplemented from the topic modeling results. Both healthcare providers and other medical organizations find patient experiences vital. Therefore, they typically collect patient experiences using questionnaires after hospital visits. Among those surveys, the HCAHPS is one of the most widely-adopted surveys. The HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems) survey is the first national, standardized, publicly reported survey of patients' perspectives of hospital care. The survey is conducted after patients are discharged, through mail, phone, interactive voice response, and mixed ways. Over 4,000 hospitals participate in HCAHPS, and over 3.0 million patients complete the survey each year. The 19 core questions cover eight aspects of patients' hospital experiences, including doctor, nurse, hospital staff, hospital room, medicines, discharge information, overall rating, and recommendation intention.[5] Because this essay investigates specific aspects of patient experience, we exclude the two aspects that reflect overall evaluations (overall rating and recommendation intention). We keep the other six concepts in HCAHPS, namely “Doctor,” “Nurse,” “Staff,” “Hospital Room,” “Discharge Information,” and “Medicines.”

[5] The detailed information on HCAHPS is available on the official website, www.hcahpsonline.org.

Besides the HCAHPS survey, many previous studies investigate patients' experience classification. For example, Bowling et al. (2012) review 211 papers, search five major electronic datasets, and survey 833 patients. They summarize various categories of patients' pre-visit expectations and post-visit experiences. We further inherit nine categories from their study, including “Appointment,” “Diagnosis,” “Facility,” “Medical Test,” “Referral,” “Recovery Information,” “Symptom and Disease,” “Timeliness,” and “Treatment.”

As online healthcare reviews have gained popularity in recent years, many studies extract latent topics from review contents with topic modeling approaches (e.g., Saifee et al., 2019; Xu et al., 2021). Researchers believe online healthcare reviews may contain more topics than offline surveys. Ranard et al. (2016) compare the latent topics extracted by the topic modeling approach with the HCAHPS survey domains. They find that while the latent topics included in online reviews cover most of the HCAHPS domains, online reviews contain 12 additional topics not included in HCAHPS.

Topic Modeling

We apply the topic modeling method to online healthcare reviews to verify the identified domains and unveil additional topics. Latent Dirichlet allocation (LDA) is one of the most popular topic modeling approaches (Blei et al., 2003). It is an unsupervised method that can automatically discover the latent topics embedded in documents without any labeled data. LDA generates a pre-defined number of latent topics according to terms' co-occurrence. Notably, traditional topic model-based ABSA methods have two main limitations (Nguyen et al., 2019; Schouten & Frasincar, 2015). First, they commonly require large datasets to reach effective results. Second, some of the extracted topics may not be relevant. However, unlike the traditional topic model-based ABSA methods that only leverage topic modeling approaches, such as LDA, to identify latent topics, we use the topic modeling approach as a supplement and validation to the pre-defined category list. Therefore, our framework avoids being influenced by the two limitations. First, once the pre-defined categories and dictionaries are developed, they can be applied to data of any size. Furthermore, we do not keep all the extracted topics in the final list but screen out the irrelevant ones.

We apply the LDA implementation in MALLET (Machine Learning for Language Toolkit) (McCallum, 2002) through *Gensim* (Rehurek & Sojka, 2011) in Python. We set the number of topics of the LDA model to 20, which means the model creates 20 topics. The reasons are bi-fold. First, the model with 20 topics yields the highest coherence score on our data. Second, 20 is marginally higher than the number of current aspect categories (15). We expect the LDA latent topics can verify some of the existing categories and provide other crucial topics. The outputs of the LDA model are 20 topics and their corresponding keyword lists. Table 3.1 stores the top ten keywords for the 20 topics.

Next, we manually interpret the keywords and give each topic an appropriate name. The topic names are provided in Table 3.1. For example, Topic 2 corresponds to insurance and billing with terms such as "bill," "insurance," "charge," and "money." So, we name it "Insurance and Billing." Nearly all patients have to deal with insurance and billing issues for their hospital visits. Therefore, the quality of this aspect, such as payment procedure and insurance coverage, should play a vital role in patients’ experience. Topic 13 has keywords related to the dietary quality of a hospital, such as “food,” “cafeteria," "meal," and "lunch." Thus, this topic is named "Food." Dietary quality, such as accessibility and variety, could be an important service aspect of a hospital for both patients and their companions. Topic 19 describes deliveries in obstetrics with the keywords such as "baby," "deliver," "delivery," and "child." Therefore, we name it

"Childbirth." Delivery is different from other medical procedures and treatments since pregnancy cannot be classified as one disease or condition. The purpose of a hospital visit for a baby delivery is essentially different from a disease or condition. Therefore, we decide to have a separate category for "Childbirth."

Table 3. 1 LDA Topic Modeling Classifications

Topic	Keywords	Topic Name
1	doctor, test, result, order, visit, treatment, problem, diagnosis, show, lab	Diagnosis
2	bill, pay, insurance, charge, receive, send, service, money, cost, visit	Insurance and Billing
3	pain, back, arm, ray, leave, fall, break, leg, hurt, home	Symptom and Disease
4	back, home, check, send, leave, discharge, bring, start, sick, fine	Discharge Information
5	call, appointment, time, phone, answer, back, speak, information, talk, number	Appointment
6	nurse, time, feel, question, check, experience, explain, understand, make sure, talk	Nurse
7	medical, facility, review, year, health, physician, system, case, include, resident	Other
8	place, doctor, time, emergency, run, avoid, joke, awful, sick, top	Other
9	visit, area, friend, time, close, parking, small, facility, staff, walk	Other
10	room, nurse, bed, floor, night, clean, move, bathroom, dirty, chair	Hospital Room
11	patient, treat, care, treatment, employee, respect, compassion, understand, attitude, poor	Treatment
12	walk, leave, back, talk, hand, hear, put, door, start, head	Other
13	care, food, stay, staff, cafeteria, receive, meal, service, eat, lunch	Food
14	staff, care, experience, service, quick, pleasant, positive, fast, recommend, visit	Other
15	surgery, day, husband, procedure, wife, experience, surgeon, time, recovery, perform	Surgery
16	hour, wait, minute, sit, time, waiting room, long, check, triage, back	Timeliness
17	refuse, state, write, medication, lie, sign, wrong, report, request, happen	Other
18	life, die, day, heart, save, week, year, infection, due, lose	Other
19	nurse, baby, daughter, experience, son, child, deliver, time, delivery, room	Childbirth
20	nurse, family, mom, mother, care, day, admit, dad, family member, transfer	Family Member

“Insurance and Billing,” “Food,” and “Childbirth” are the additional aspect categories identified from the topic modeling results. Now, we have all 18 aspect categories from existing classifications and the LDA topic modeling approach. As Table 3.1 shows, other than the new latent topics, the latent topics are largely consistent with the existing aspect categories. The eight consistent topics are: "Diagnosis," "Symptom and Disease," "Discharge Information," "Appointment," "Nurse," "Hospital Room," "Treatment," and "Timeliness." The aligned results indicate the appropriateness of our aspect categories. The latent topics "Surgery" and "Family Members" are not kept in our final categories. First, "Surgery" is merged with the "Treatment" category. Second, the keywords in the "Family Member" topic are general terms (i.e., "mom," "daughter," and "dad"). Therefore, if we use those general terms as keywords, the accuracy of the results will be influenced. Notably, seven topics contain mixed and less informative keywords. Therefore, we name them “Other” and exclude them from the final aspect category list.

Development of Dictionaries

After defining the 18 aspect categories, we develop one dictionary for each category. The creation of dictionaries consists of three steps. We first propose several core terms for each category that correlate most to the context. Next, we extend the dictionaries by including the core terms' synonyms and relevant terms. Finally, to connect the dictionaries to our corpus, we enrich them with the matched consumer healthcare vocabulary (CHV) terms. The detailed steps are introduced in the following sections.

Identify Core Terms

We start by determining the core terms for each category. Core terms are vocabularies that can represent their corresponding categories. For example, the five core terms for the “Doctor” category are: “doctor,” “physician,” “hospitalist,” “surgeon,” and “clinician.” They are different names for doctors and are used frequently in patients’ communications.

Include Synonyms and Relevant Terms

Starting from the core terms, we extend the dictionaries first by including the core terms' synonyms and relevant terms. For the "Doctor" category, the synonyms are the terms that have a meaning of doctors. We further include terms such as "specialist," "therapist," "practitioner," and "healer." Then, relevant terms are those in the same category as the core terms. In the "Doctor" category, the relevant terms can be the titles for different specialties, such as “pediatrician,” “neurologist,” “radiologist,” “urologist,” “obstetrician,” and “dermatologist.”

Add Matched CHV Terms

The major limitation of the vocabulary-based approach is that the pre-defined dictionaries and the review corpus may not link closely. Specifically, there is no guarantee that the terms in the pre-defined dictionaries will be present in the review contents (Al-Ghuribi & Noah, 2019; Gaputo et al., 2017). To overcome this drawback, we further enrich the matched CHV terms in the reviews in the dictionaries.

CHV is designed to complement the existing Unified Medical Language System (UMLS) knowledge. Its terms focus on expressions and concepts that are employed by health-related communications from or to consumers. To extract medical terms in our reviews, we match the textual contents with CHV. After the matching process, we have a list of CHV terms that occurred in the reviews. However, CHV does not classify medical terms into different categories. Therefore, we manually screen the matched terms and group the relevant ones into corresponding categories. As a result, our dictionaries can appropriately connect with the reviews in our data. For example, the following terms are the matched CHV terms for the “Doctor” category: “family doctor,” “primary care physician,” “speech therapist,” and “plastic surgeon.” The descriptions and representative keywords for the 18 pre-defined aspect categories’ dictionaries are summarized in Table 3.2.

Table 3. 2 Dictionaries for the Aspect Categories

Aspect Categories	Dictionary Descriptions	Representative Keywords
Appointment	The dictionary contains nouns that relate to hospital appointments or scheduling processes.	appointment, schedule, scheduling, assignation, arrangement
Childbirth	The dictionary has nouns that mean childbirth, labor, or delivery. It also contains the terms for the relevant conditions, such as "abnormal pregnancy" and "postpartum depression," and procedures, such as "C-section."	childbirth, delivery, pregnancy, trimester, parturition
Facility	The dictionary includes nouns for hospital buildings and facilities.	building, facility, garage, dispenser, atm
Diagnosis	The dictionary contains nouns that relate to doctors' diagnosis processes.	diagnosis, touch_look_compare, lab_report, inspection, medical_sign
Discharge Information	The dictionary has nous that represent information relevant to a patient's health issue provided by hospitals at the time of the patient's discharge.	discharge_information, discharge_instruction, follow_up treatment, discharge_conversation, care_at_home
Doctor	The dictionary includes nouns for different types of doctors.	doctor, physician, surgeon, neurologist, family_doctor
Food	The dictionary contains nouns that are relevant to food or meals.	food, meal, cafeteria, restaurant, eatery
Insurance and Billing	The dictionary has nouns that are relevant to insurance and billing information.	insurance, doctor_bill, payment, insurance_coverage, health_plan
Medical Test	The dictionary includes nouns and abbreviations for common medical tests.	medical_test, lab_test, blood_test, biopsy, MRI
Medicine	The dictionary contains Nouns for common medicines.	medicine, capsule, advil, amoxicillin, penicillin
Nurse	The dictionary has nouns for different types of nurses.	nurse, nursemaid, regeistered_nurse, practical_nurse,CRNA
Recovery	The dictionary includes nouns relate to recovery processes.	recovery, healing, amelioration, on_the_mend, out_of_the_woods
Referral	The dictionary contains nouns that refer to the process in which healthcare providers at lower levels of the health system seek	referral, reference, referral_hospital,

	assistance from high-level providers to take over responsibility for a particular episode of a clinical condition in a patient (Al-Mazrou et al., 1990).	referral_system, referral_service
Hospital Room	The dictionary includes nouns for different hospital rooms.	hospital_room, critical_room, emergency_unit, labor_room, patient_stay
Staff	The dictionary contains nouns for different medical staff except for doctors and nurses.	hospital_staff, reception, front_desk, administrator, transporter
Symptom and Disease	The dictionary has nouns for common symptoms and diseases.	symptom, ache, discomfort, sepsis, afib
Timeliness	The dictionary includes nouns relate to the waiting time in hospitals.	waiting_time, wait_time, timeliness, timeline, waiting_room
Treatment	The dictionary contains nouns for common treatments, surgeries, and procedures.	treatment, surgery, therapy, IV, transplant

Data Preparation

The data preparation consists of two parts. We first clean the review contents using standard text cleansing processes, including stop word removal, tokenization, lemmatization, bigram creation, and trigram creation. We then replace the phrases (more than one word) in the dictionaries with n-grams using Python. Finally, we match and replace the phrases with n-grams in the reviews. The details are discussed in the following sections.

Standard Text Preparation Steps

Some text preparation steps are common for various tasks. We first perform stop word removal, tokenization, lemmatization, bigram creation, and trigram creation in this essay using Python. Stop words are generally the most common and less informative words. For example, "the," "and," "so," and "a." By removing these words, we eliminate the less informative words from our text so that we can focus more on the important terms. In other words, the stop words provide little information for understanding text. Additionally, removing stop words can reduce the corpus size and thus accelerate the program process speed. The initial stop word set is downloaded from Python's natural language toolkit (NLTK) package. We then modify the stop word list by excluding negation words "no" and "not." It is because negation words overturn sentiments. Then, to ensure that the terms in the dictionaries are kept in the final corpus, we exclude any overlapped terms from the stop word list and dictionaries.

Tokenization is the process of breaking text documents into units of meaning, which usually are words. However, we conduct the ABSA at a sentence level. Thus, we tokenize reviews into sentences using the "sent_tokenize" function in the "NLTK" package. Lemmatization is a process that transfers the inflected forms of a word to its base form or dictionary form, known as the word's lemma. For example, the lemma for "loves," "loved," and "loving" is "love." We perform the lemmatization with the "spaCy" package in Python. Finally, we create bigrams and trigrams. Bigrams are two words frequently occurring together in the document, and trigrams are three words that frequently occur together. We implement this step with the "Gensim" package's "Phrases" model.

Manually Create N-grams

The automatic n-gram creations are based on the co-occurrences of multiple words. Creating n-grams helps parse reviews in a better way. However, the phrases in the dictionaries may not frequently occur in reviews. For example, some medical tests, such as "upper gastrointestinal endoscopy," "transvaginal ultrasound," and "muscle biopsy," are scarce in reviews. Meanwhile, they are keywords in the "Medical Test" dictionary. Therefore, if we rely solely on programs to create n-grams automatically, many of the dictionary phrases will not be identified and transferred.

One of the most crucial tasks for an accurate ABSA is identifying keywords and sentiment terms using part-of-speech (POS) tags. A phrase is supposed to be identified and tagged as a noun, as a whole. Failing to replace the phrase with an n-gram may cause inaccurate

POS tags, which would lead to biased ABSA results. Furthermore, such a bias can also lead to incorrect word dependencies. To better illustrate the necessity of creating n-grams for all phrases in the dictionaries, let us consider the following two sentences:

Sentence A: The transvaginal ultrasound is quick and painless.

Sentence B: The transvaginal_ultrasound is quick and painless.

As can be seen, the two sentences are identical, except sentence A contains the original phrase for "transvaginal ultrasound," and sentence B uses its bigram. The followings are the tagged results using spaCy's "Token.pos_" attributes in Python.

Tagged Sentence A: The transvaginal (ADJ) ultrasound (NOUN) is quick and painless.

Tagged Sentence B: The transvaginal_ultrasound (NOUN) is quick and painless.

The program identifies the *transvaginal* as an adjective and the *ultrasound* as a noun for sentence A. If this tagged sentence is passed to the ABSA function, *transvaginal* will be treated as the descriptive term for the keyword *ultrasound,* which is a mistake. If we replace the phrase with its bigram as sentence B does, then *"transvaginal_ultrasound"* will be recognized as a noun, as a whole. Then, *quick* and *painless* will be correctly identified as the descriptive terms for the bigram.

Therefore, we write two Python functions to create n-grams for all phrases with more than one word in the pre-defined dictionaries and replace all matched phrases in the review contents with their corresponding n-grams. As a result, all phrases in the dictionaries and corpus are replaced with n-grams.

Sentiment Analysis

After having the cleaned reviews and dictionaries, we can perform aspect-based sentiment analysis. This section has three steps. We first extract keyword-description pairs from reviews using the POS tags and word dependencies. Second, we match the keywords with the terms in the dictionaries and extract the matched keywords. Finally, we calculate the sentiments of all matched keywords and compute the average sentiments for the different aspect categories.

Keyword-description Pairs Extraction

We identify keywords and descriptions according to the POS tags and word dependencies using the English package "en_core_web_sm" in the "spaCy" package in Python. "spaCy" is a free and open-source library for Natural Language Processing in Python. Its POS tagging function uses a trained pipeline and statistical models to predict tags or labels of terms in documents[6]. For example, each word in a sentence can be tagged as a verb (VERB), noun (NOUN), adjective (ADJ), or adverb (ADV). The unit of the analysis is at the sentence level. The candidate words for keywords are usually nouns or noun phrases (Moghaddam & Ester, 2010; Hu & Liu, 2004; Eirinake et al., 2012; Mubarok et al., 2017). Therefore, we extract nouns from sentences and keep them as keywords. Then, the terms that describe the keywords are typically adjectives and adverbs in the same sentence. We summarize five patterns for descriptive terms. 1. Only one adjective, such as "friendly." 2. One negation and one adjective, such as “not friendly.” 3. One adverb and one adjective, such as “very friendly.” 4. One negation, one adverb, and one adjective, such as “not very friendly.” 5. One adverb, one negation, and one adjective, such as “very not friendly.” Next, we screen the extractions from POS tags by filtering the dependencies of the keywords and descriptive terms using “spaCy.” The dependencies have to be one of the certain relations, such as adjectival modifier (*amod)*, adverb modifier (*advmod),* and nominal subject (*nsubj),* to be kept in the final list. After this step, the keywords and their corresponding

---

[6] https://spacy.io/usage/linguistic-features

descriptive terms in each sentence are retrieved and stored in a list. For example, the original sentence is: "*The nurse is very nice and friendly.*" When we pass this sentence to the function, the output will be a list of lists, where the first sub-list stores the keywords and the second sub-list stores the descriptive terms. The output will be like: *"[["nurse"], ["very nice", "friendly"]].*"

Keywords Match and Sentiment Calculation

After retrieving the keyword-description pairs, we pass the pairs onto the next function, where we match the keywords with our pre-defined vocabularies. To capture the misspelled words in the reviews, we adopt the fuzzy match function instead of the exact match using the "fuzz" function from the "fuzzywuzzy" package in Python. Ignoring all of the keywords that reviewers incorrectly write negatively affects the ABSA process, but this problem can be eliminated by using the "fuzzywuzzy" package due to its ability to calculate the similarity values between two strings based on Levenshtein Distance. The fuzzy ratio is the measure of similarity. The higher the ratio, the more similar the two strings are. It ranges from 0 to 100, where 100 refers to an exact match. In our program, we set the fuzzy ratio threshold to 95. In other words, if two strings' fuzzy ratio is higher than 95, they will be recognized as the same term.

Finally, we calculate the sentiment polarities for the matched keywords' descriptions. Unlike many previous studies that only classify an aspect as either positive, negative, or neutral (e.g., Al-Ghuribi et al., 2020; Chen & Yao, 2010; Hu & Liu, 2004), we calculate numeric polarities using the "TextBlob" package in Python. The numeric polarities help us gain more insights when we include them in regression models. After having the sentiment scores for the matched keywords, we calculate the aspect categories' average sentiment scores by averaging the polarities for all the keywords that belong to the same category. We then further integrate the aspect categories' sentiments into regression models and investigate their effects on hospitals' online review ratings.

Data

The data collection process starts by determining the target hospitals. We select from the hospitals that enroll in the Centers for Medicare and Medicaid Services (CMS) *Hospital Compare* program. Hospital Compare includes information for over 4,000 acute care and critical access hospitals. Its public datasets contain hospitals' demographic information, which we use in our econometric models as control variables. We collect and merge the hospital datasets in five years (2016-2020). The Hospital Compare program is voluntary. Thus, the participating hospitals and available information vary over the years. Therefore, when we merge the datasets from 2016 to 2020 into one table, we only reserve the hospitals that have complete demographic information over the five years. Next, we use the Python program to collect online reviews for the hospitals on Yelp.com. Yelp is an online platform where users can share their evaluations online (ratings and textual reviews) of businesses such as restaurants, hotels, and hospitals. It is one of the most widely adopted commercial online review websites in the United States for hospital ratings (Bardach et al., 2013). Some hospitals have no reviews between 2016 and 2020. After removing those hospitals, the final dataset contains 34,129 reviews for 686 hospitals from 49 states across the US in the five years. The number of reviews for each hospital ranges from 2 to 392. Finally, after the ABSA analysis, 4,697 reviews contain no aspect. Thus, we remove them and get our final dataset, which has 29,432 reviews.

Variables

The dependent variable is online rating, the numeric rating a reviewer gives to a hospital on a 5-point scale on Yelp.com, indicating her overall evaluation of the hospital.

The focal variables are 18 aspect categories' average sentiment scores. They are extracted from the ABSA framework introduced above and are on the same scale from -1 to 1, meaning extremely negative to extremely positive. If a review does not contain certain categories, the sentiment scores will be 0.

We also control for a comprehensive set of control variables for reviewer, review, review content, and hospital information. Prior online review studies have largely recognized the importance of reviewers' characteristics for search or experience products or services. For example, by examining the restaurant reviews, Zhang and Liu (2019) find that a reviewer's number of friends negatively affects her review ratings. However, the impact of reviewers' characteristics on healthcare reviews has not been investigated. Following the previous online review literature, we control for reviewer-related variables, including reviewer number of friends (Banerjee et al., 2017), number of reviews (Cheng & Ho, 2015), number of badges (Zhu et al., 2014), and number of followers (Luo et al., 2021). In addition to the reviewer information, review and review content information have consistently been recognized as important factors in the online review domain (Ghose & Ipeirotis, 2010; Yin et al., 2014). Therefore, we further control for review length, subjectivity, and fog index.

We also use the following hospital control variables in our models: number of beds, emergency department volume, hospital type, hospital ownership, and hospital state. Table 3.3 summarizes the information for all the variables used in this essay.

Table 3. 3 Variable Summary

Group	Variable Name	Description	Summary Statistics: Mean (SD.)
Dependent Variables	Online Rating	The online rating is rated by reviewers for hospitals on Yelp.com.	2.587 (1.808)

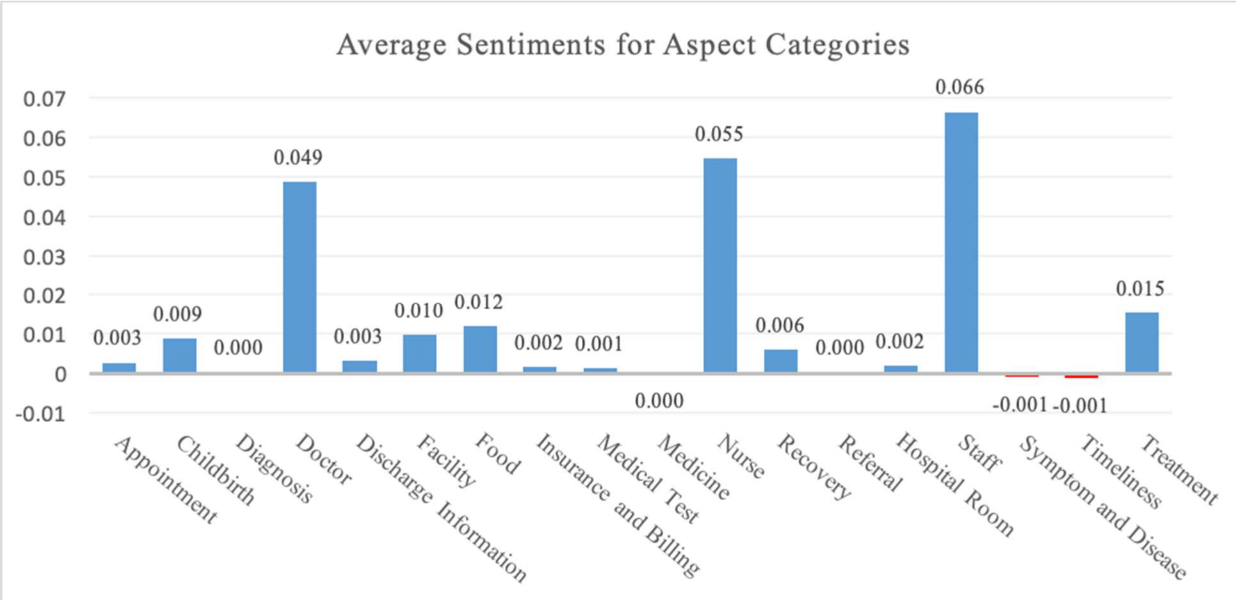
Focal Variables: Aspect Categories' Average Sentiment Scores	Appointment Sentiment	The average sentiment score for all matched keywords, which belong to the "Appointment" aspect category ([-1,1] = [negative, positive]).	0.003 (0.061)
	Childbirth Sentiment	The average sentiment score for all matched keywords, which belong to the "Childbirth" aspect category ([-1,1] = [negative, positive]).	0.009 (0.081)
	Facility Sentiment	The average sentiment score for all matched keywords, which belong to the "Facility" aspect category ([-1,1] = [negative, positive]).	0.010 (0.109)
	Diagnosis Sentiment	The average sentiment score for all matched keywords, which belong to the "Diagnosis" aspect category ([-1,1] = [negative, positive]).	0.0004 (0.067)
	Discharge Information Sentiment	The average sentiment score for all matched keywords, which belong to the "Discharge Information" aspect category ([-1,1] = [negative, positive]).	0.003 (0.072)
	Doctor Sentiment	The average sentiment score for all matched keywords, which belong to the "Doctor" aspect category ([-1,1] = [negative, positive]).	0.049 (0.214)
	Food Sentiment	The average sentiment score for all matched keywords, which belong to the "Food" aspect category ([-1,1] = [negative, positive]).	0.012 (0.119)
	Insurance and Billing Sentiment	The average sentiment score for all matched keywords, which belong to the "Insurance and Billing" aspect category ([-1,1] = [negative, positive]).	0.002 (0.081)
	Medical Test Sentiment	The average sentiment score for all matched keywords, which belong to the "Medical Test" aspect category ([-1,1] = [negative, positive]).	0.001 (0.045)
	Medicine Sentiment	The average sentiment score for all matched keywords, which belong to the "Medicine" aspect category ([-1,1] = [negative, positive]).	-0.0001 (0.085)
	Nurse Sentiment	The average sentiment score for all matched keywords, which belong to the "Nurse" aspect category ([-1,1] = [negative, positive]).	0.055 (0.222)
	Recovery Sentiment	The average sentiment score for all matched keywords, which belong to the "Recovery" aspect category ([-1,1] = [negative, positive]).	0.006 (0.067)
	Referral Sentiment	The average sentiment score for all matched keywords, which belong to the "Referral" aspect category ([-1,1] = [negative, positive]).	0.0002 (0.014)
	Hospital Room Sentiment	The average sentiment score for all matched keywords, which belong to the "Hospital Room" aspect category ([-1,1] = [negative, positive]).	0.002 (0.044)
	Staff Sentiment	The average sentiment score for all matched keywords, which belong to the "Staff" aspect category ([-1,1] = [negative, positive]).	0.066 (0.240)
	Symptom and Disease Sentiment	The average sentiment score for all matched keywords, which belong to the "Symptom and	-0.001 (0.145)

		Disease" aspect category ([-1,1] = [negative, positive]).	
	Timeliness Sentiment	The average sentiment score for all matched keywords, which belong to the "Timeliness" aspect category ([-1,1] = [negative, positive]).	-0.001 (0.074)
	Treatment Sentiment	The average sentiment score for all matched keywords, which belong to the "Treatment" aspect category ([-1,1] = [negative, positive]).	0.015 (0.149)
Control Variables: Reviewer Information	Reviewer Number of Friends	The number of friends of a reviewer.	68.985 (248.262)
	Reviewer Number of Reviews	The number of past reviews posted by a reviewer.	60.008 (262.373)
	Reviewer Number of Badges	The number of badges a reviewer has received.	0.422 (1.546)
	Reviewer Number of Followers	The number of followers of a reviewer.	3.077 (31.012)
Control Variables: Review Information	Review Length	The number of words of a review.	185.953 (169.382)
	Review Subjectivity	Subjectivity score of a review ([0,1]=[objective, subjective])	0.526 (0.143)
	Review Fog Index	Fog index of a review	16.151 (20.275)
Control Variables: Hospital Information	Number of Beds	Hospital number of beds.	393.875 (349.710)
	EDV	Hospital emergency department volume. Four levels: low, medium, high, and very high.	Categorical Variable
	Hospital Type	Hospital type. Two types: acute care and critical access hospitals.	Categorical Variable
	Hospital Ownership	Hospital ownership.	Categorical Variable
	Hospital State	The state where the hospital is located.	Categorical Variable

Descriptive Analysis

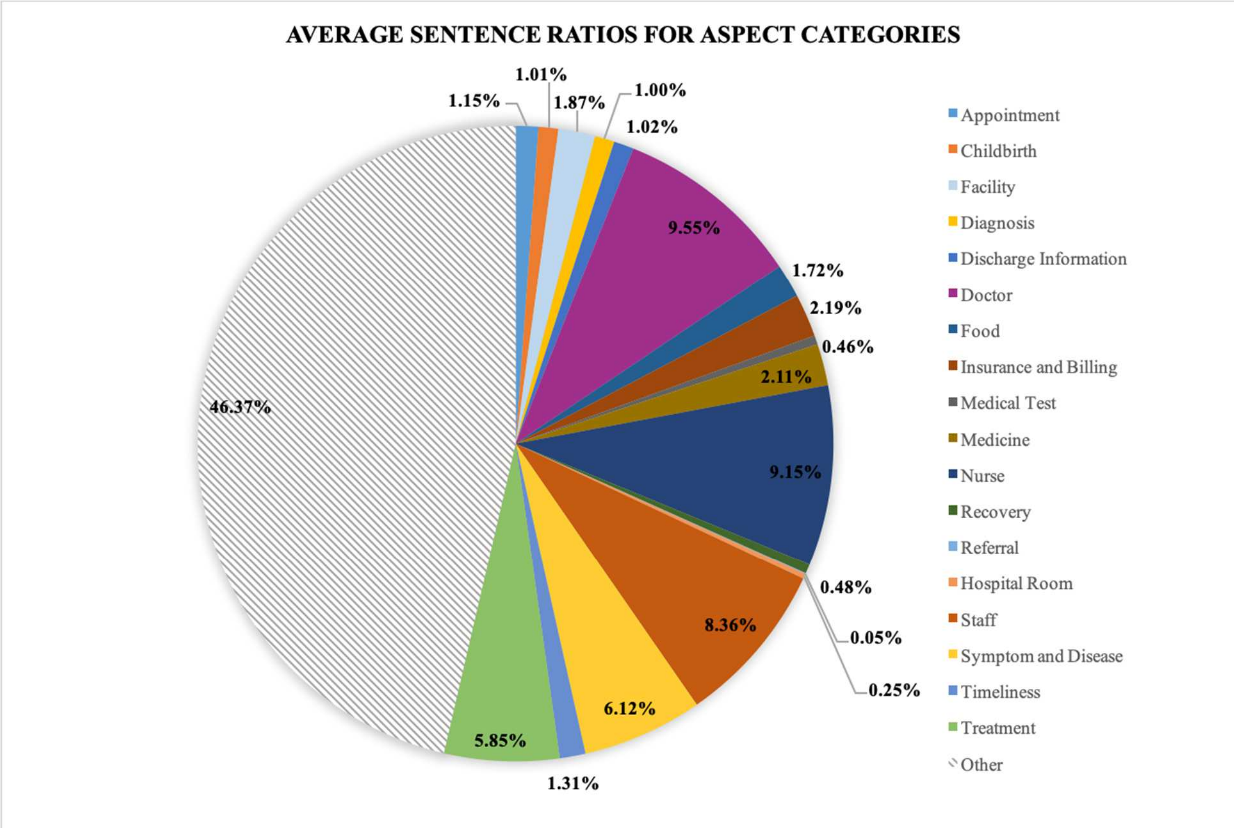
After extracting aspect categories' sentiments, we create several visualizations to have a deeper insight into them and demonstrate the usefulness of investigating the sentiments for different aspect categories in online healthcare reviews. Figure 3.2 is a histogram that shows the average sentiments for the 18 aspect categories. As can be seen, based on the reviews in our dataset, "Staff" has the highest average sentiments, followed by "Nurse" and "Doctor." It means that reviewers are most satisfied with these categories on average.

Figure 3. 2 Average Sentiments for Aspect Categories



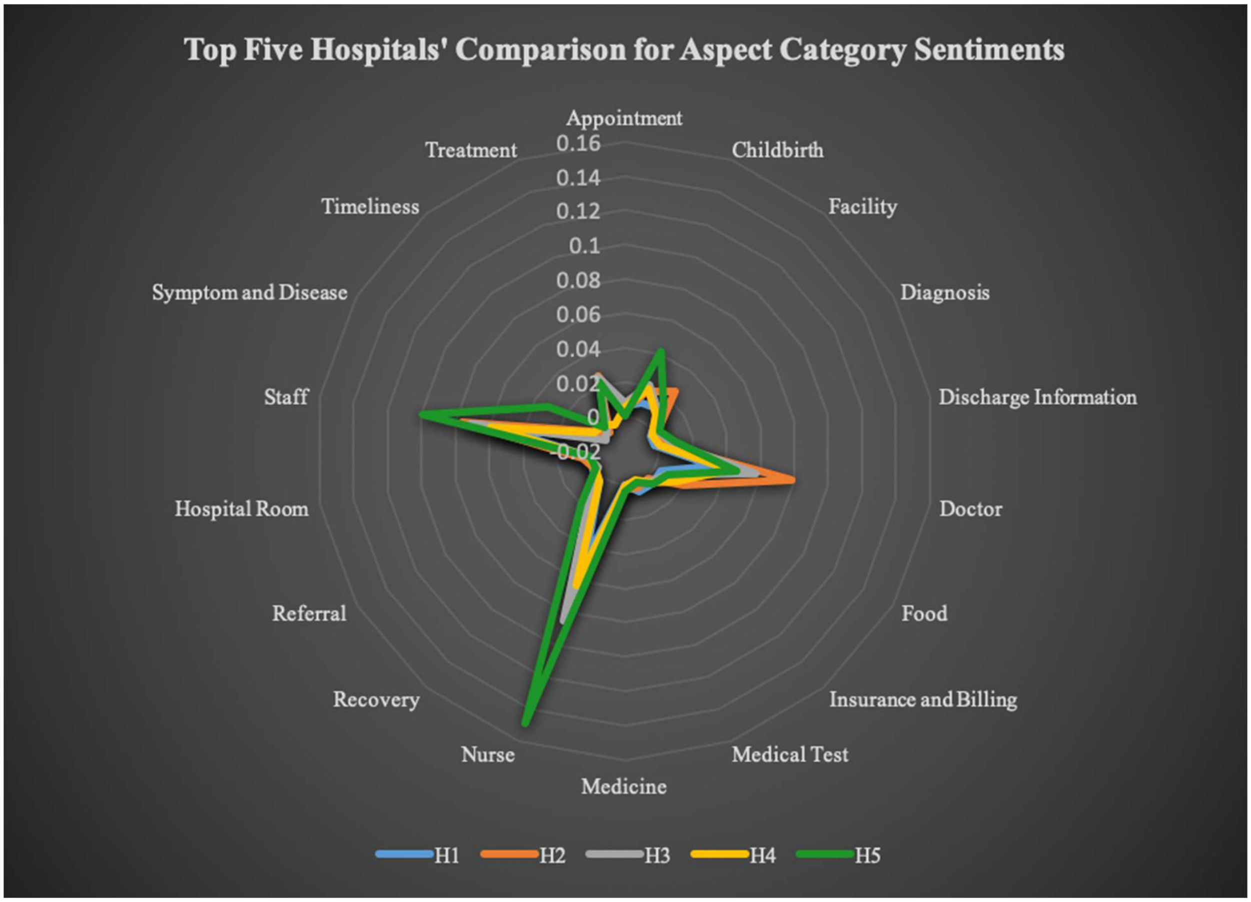
Other than the sentiments, the aspect categories' average proportions in a review are also important. We measure the proportions by the sentence ratios since we conduct the ABSA analysis at the sentence level. The sentence ratios are calculated as the ratios between the number of sentences containing a certain aspect category and the reviews' total number of sentences. For example, if a ten-sentence review has two sentences for the "Doctor" category, the sentence ratio for the category will be 0.2. Figure 3.3 shows the average sentence ratios for the 18 aspect categories. Figure 3.3 shows that besides the "Other" information, the reviews contain the "Doctor" the most, with an average ratio of 9.55%. In other words, a reviewer mentions an average of 9.55% of doctors' information in a review. The second most-mentioned category is "Nurse" with a ratio of 9.15%, followed by "Staff" (8.36%)

Figure 3. 3 Average Sentence Ratios for Aspect Categories



Finally, except for investigating the average situations of the identified categories, we can also compare the reputations of individual hospitals from the 18 different perspectives. Figure 3.4 illustrates an example that compares the reputations of the top five hospitals that accumulate the most reviews in our datasets. The numbers of reviews for those hospitals in our dataset are 336, 315, 296, 281, and 275 in descending order. The radar chart in Figure 3.4 unveils some underlying patterns. It contains 18 spokes representing 18 aspect categories. The spoke length indicates the sentiment polarity value. The highest polarity across all data points. Therefore, it is clear that the hospital "H5" outperforms the other four hospitals in many categories, such as "Nurse," "Staff," and "Childbirth." However, in the categories like "Doctor" and "Diagnosis," hospital "H2" performs the best.

Figure 3. 4 Top Five Hospitals' Comparison for Aspect Category Sentiments



Data Analysis

After obtaining the 18 aspect categories' average sentiment scores, we can now study their statistical effects on hospitals' online ratings as the final step of our proposed framework. The model is shown below:

$$
\begin{aligned}
&Online\ Rating_{rj} \\
&\quad = \beta_0 + \beta_1 Appointment_{rj} + \beta_2 \mathrm{Childbirth_{rj}} + \beta_3 \mathrm{Diagnosis_{rj}} \\
&\quad + \beta_4 \mathrm{Discharge\ Information_{rj}} + \beta_5 \mathrm{Doctor_{rj}} + \beta_6 \mathrm{Facility_{rj}} + \beta_7 Food_{rj} \\
&\quad + \beta_8 Insurance\ and\ Billing_{rj} + \beta_9 Medical\ Test_{rj} + \beta_{10} Medicine_{rj} \\
&\quad + \beta_{11} \mathrm{Nurse_{rj}} + \beta_{12} Recovery_{rj} + \beta_{13} Referral_{rj} + \beta_{14} Hospital\ Room_{rj} \\
&\quad + \beta_{15} Staff_{rj} + \beta_{16} Symptom\ and\ Disease_{rj} + \beta_{17} Timeliness_{rj} \\
&\quad + \beta_{18} Treatment_{rj} + \beta_{19} Number\ of\ Friend_{rj} + \beta_{20} Number\ of\ Reviews_{rj} \\
&\quad + \beta_{21} Number\ of\ Badges_{rj} + \beta_{22} \mathrm{Number\ of\ Followers_{rj}} \\
&\quad + \beta_{23} Review\ Length_{rj} + \beta_{24} Subjectivity_{rj} + \beta_{25} \mathrm{Fog\ Index_{rj}} \\
&\quad + \beta_{26} NumberofBeds_{rj} + \beta_{27} EDV\_\ Medium_{rj} + \beta_{28} EDV\_High_{rj} \\
&\quad + \beta_{29} EDV\_VeryHigh_{rj} + \beta_{30} HospitalType_{rj} + \alpha_{\mathrm{j}} + \theta_j + \varepsilon_{rj} \qquad (1)
\end{aligned}
$$

The dependent variable of equation (1) is the online rating. It is the numeric rating rated by reviewers for hospitals on Yelp on a 5-point scale. β_1 to β_{18} measures the effects of the 18 aspect categories' average sentiments in review r for hospital j. β_{19} to β_{30} are the coefficients of the control variables. α_{j} and θ_j denote the fixed effects of hospital ownerships and state correspondingly.

Empirical Results

We expect that the sentiments of the individual aspect categories should influence hospitals' online ratings. Therefore, we build the linear regression models to unveil the aspect categories' comparative importance. The results are stored in Table 3.4. Model 1 is the baseline model, which has only the control variables. Model 2 further includes the average sentiment scores of 18 aspect categories and shows the original coefficients. To compare the relative importance of the aspect categories, we also report the standardized coefficients obtained by normalizing the variables. The standardized results can be seen in Model 3.

Table 3. 4 Analysis Results.

	Model 1 (Baseline)	Model 2 (Original)	Model 3 (Standardized)
(Intercept)	2.289 (0.4538) ***	2.386 (0.392) ***	2.706 (0.392) ***
Appointment		0.5979 (0.1418) ***	0.0363 (0.0086) ***
Childbirth		1.443 (0.1072) ***	0.1169 (0.0087) ***
Diagnosis		0.1711 (0.1303)	0.0115 (0.0087)
Discharge Information		0.6407 (0.1201) ***	0.046 (0.0086) ***
Doctor		0.9343 (0.0431) ***	0.1997 (0.0092) ***
Facility		0.9714 (0.0796) ***	0.1058 (0.0087) ***
Food		1.232 (0.0731) ***	0.1465 (0.0087) ***
Hospital Room		0.6093 (0.1962) **	0.0268 (0.0086) **
Insurance and Billing		0.0844 (0.1067)	0.0068 (0.0086)
Medical Test		0.6102 (0.1925) **	0.0273 (0.0086) **
Medicine		-0.0647 (0.1051)	-0.0055 (0.0089)
Nurse		1.468 (0.0417) ***	0.3259 (0.0093) ***
Recovery		1.296 (0.1286) ***	0.0871 (0.0086) ***
Referral		0.1437 (0.5948)	0.0021 (0.0086)
Staff		1.899 (0.0375) ***	0.4555 (0.009) ***
Symptom and Disease		0.7273 (0.0613) ***	0.1052 (0.0089) ***
Timeliness		0.6286 (0.1166) ***	0.0464 (0.0086) ***
Treatment		1.247 (0.0605) ***	0.1853 (0.009) ***
Number of Friends	0.0002 (0.0001) ***	0.0002 (0) ***	0.0002 (0) ***
Number of Reviews	0.0001 (0.0001)	0.0001 (0)	0.0001 (0)
Number of Badges	0.2113 (0.008) ***	0.1545 (0.0069) ***	0.1545 (0.0069) ***
Number of Followers	-0.0009 (0.0005)	-0.0011 (0.0004) **	-0.0011 (0.0004) **
Review Length	-0.002 (0.0001) ***	-0.0018 (0.0001) ***	-0.0018 (0.0001) ***
Subjectivity	1.937 (0.0705) ***	1.247 (0.0614) ***	1.247 (0.0614) ***
Fog Index	-0.0012 (0.0005) *	-0.0009 (0.0004) *	-0.0009 (0.0004) *
Number of Beds	0 (0)	0 (0)	0 (0)
EDV: High	-0.0745 (0.0393)	-0.0789 (0.034) *	-0.0789 (0.034) *
EDV: Medium	0.0523 (0.0373)	0.0162 (0.0322)	0.0162 (0.0322)
EDV: Very High	-0.0843 (0.0379) *	-0.063 (0.0327)	-0.063 (0.0327)
Hospital Type: Critical Access Hospital	0.2402 (0.1058) *	0.1036 (0.0914)	0.1036 (0.0914)
Hospital Ownership	Yes	Yes	Yes
Hospital State	Yes	Yes	Yes
R Square	0.1136	0.3387	0.3387

*** p<0.001, ** p<0.01, * p<0.05

As we can see in Table 3.4, most of the 18 aspect categories have positive and significant effects on overall online review ratings, except for “Diagnosis,” “Insurance and Billing,” “and “Medicine.” Especially, the significant effects of “Childbirth” and “Food” support the necessity of supplementing aspect categories using the Topic Modeling approach. The significant increase in R square values from Model 1 to Model 2 indicates that including the aspect categories’ sentiment scores dramatically increases the model's fit. Model 3 contains the standardized coefficients, which reveal the relative importance of the aspect categories from the coefficients' magnitudes. "Staff" ($\alpha = 0.4555, p < 0.001$), "Nurse" ($\alpha = 0.3259, p < 0.001$), and "Doctor" ($\alpha = 0.1997, p < 0.001$) are the three most crucial aspect categories for hospitals' online reputations.

Discussion

As online reviews have gained importance in the patient decision process in the healthcare domain (Gao et al., 2015, Hedges & Couey, 2019) in recent years, an increasing number of studies have studied various factors that affect healthcare providers' online reputation (e.g., Gao et al., 2012, Gao et al., 2015). However, no study investigates how aspect-based sentiments analysis (ABSA) in online healthcare reviews influences the overall numeric online ratings. This essay proposes a framework that combines the ABSA technique and regression models to investigate the determinants of hospital online ratings. The proposed ABSA combines vocabulary-based, syntactic relation-based, and topic model-based techniques. We innovatively introduce the consumer healthcare vocabulary in the framework to ensure the relevance of the pre-defined vocabularies to the review corpus. We also include an additional step to replace phrases with n-grams in the vocabularies and the corpus to increase the accuracy of the part-of-speech tags and word dependencies. Our proposed method can overcome major limitations in the exiting ABSA techniques. Then, we adopt the framework to extract 18 aspect categories and calculate their corresponding sentiment scores. Finally, we include the sentiment scores in the regression models to study their effects on online ratings, which is also the first attempt to the best of our knowledge.

The results of the regression models reveal that most of the 18 aspect categories' sentiment scores positively affect numeric online ratings and the relative importance between them. Moreover, including the aspect categories' sentiment scores significantly improves the model's fit, which proves the importance of the embedded patient opinions to their overall evaluations.

Among the significant factors, “Staff,” “Nurse,” and “Doctor” are the three most important factors for online review ratings. This finding can provide valuable information for hospitals to improve their online reputations. Specifically, improving the performance of their staff, nurses, and doctors could be the best way to boost their online review ratings. In other words, patients' interactions with a hospital’s staff, nurses, and doctors affect the overall evaluations of that hospital the most. Surprisingly, the most influential category is "Staff," which includes hospital staff such as receptionists, transporters, and administrators. It indicates that compared to the evaluations of nurses and doctors, patients' opinions of other hospital staff are even more crucial in forming their overall evaluations. The importance of "Treatment" and "Food" is in the fourth and fifth place, respectively.

In contrast, “Diagnosis,” “Insurance and Billing,” “Medicine,” and "Referral" do not have statistically significant impact on review ratings. The quality of diagnosis and referral processes is hard to evaluate for patients, especially when they have limited medical knowledge. Therefore, patients' attitudes towards these two categories are insignificant in their overall evaluations. The medicine, insurance, and billing perspectives are relatively standard across hospitals, reflected by their near-zero average values. Patients commonly have neutral sentiments toward these two categories, so they cannot differentiate the performance of different hospitals; thus, they are not significant.

The results have important managerial implications. Patients can understand that the most crucial aspect categories for hospitals' overall ratings are "Staff," "Nurse," and "Doctor." Therefore, if they also value these three perspectives when they choose hospitals, the review ratings can be appropriate references. However, if patients care more about "Diagnosis," "Insurance and Billing," "Medicine," and "Referral," the review ratings may not be reasonable indicators. Instead, they should look into the review content and find independent evaluations on these aspects. As for hospitals, they can use the regression models' results as references to improve their online reputations effectively. For example, suppose a hospital plans to increase its food diversities or improve the appointment system to boost its online rating. In that case, they should prioritize the food improvement since patients value the food category more than the appointment, according to our results.

Moreover, hospitals can leverage the ABSA results to compare their performance with their major competitors' using visualizations (see, for example, Figure 3.4). We suggest that online healthcare review platforms display sentiment scores for different aspects extracted from the contents. Our empirical results suggest that different aspect categories influence overall review ratings differently. Therefore, solely providing review ratings may not comprehensively reflect reviewers' evaluations of hospitals with respect to every aspect.

Conclusions, Limitations, and Future Directions

This essay proposes a hybrid framework merging ABSA and regression models for online healthcare reviews to extract embedded aspect categories' sentiments and their influence on online ratings. We first inherit aspect categories from existing surveys and literature. We then further identify additional categories from the topic modeling results. The final list has 18 aspect categories. After having the aspect category list, we develop the dictionaries. We start with proposing core terms for each category. Next, we enrich the dictionaries by including core terms' synonyms and relative terms. Finally, to build a bridge between the dictionaries and the corpus, we further extend the dictionaries with the matched CHV terms in the reviews. Next, we perform standard text preparation steps, followed by the n-grams replacement, replacing phrases with n-grams in the dictionaries and reviews to increase the accuracies of POS tags and word dependencies. We then use the keywords in the dictionaries to locate their corresponding sentiment words and calculate the average sentiment scores.

After getting the aspect categories' sentiment scores, we include them in the regression models to investigate how different categories' sentiments affect overall numeric ratings. The analysis results show that involving the 18 aspect categories' sentiment scores significantly improves model fit. Moreover, the standardized coefficients of the sentiment scores clearly distinguish the relative importance of the 18 aspect categories. Our proposed framework and analysis results can have essential implications for patients, hospitals, and online healthcare review platforms.

This essay could be improved in the future with the following perspectives in mind. First, the proposed framework calculates sentiment scores using the standard Python package "TextBlob," but we do not develop context-sensitive sentiment analysis models. It is because of the lack of domain-specific sentiment lexica (Zunic et al., 2020). Future studies can develop such a lexica and train sentiment analysis models using machine learning techniques, such as support vector machine (e.g., Daniulaityte et al., 2016; Zhou et al., 2015), naïve Bayes (e.g., Du et al., 2017; Metwally et al., 2017), and decision trees (Greaves et al., 2013; Islam et al., 2018). Second, we measure the sentiment polarity of an aspect category by averaging the sentiment scores of all matched descriptive terms without putting weights on different terms. Future studies can try to weight average the descriptive terms of a category to generate the final polarity referring to their relative importance measured by, for instance, term frequency-inverse document frequency (TF-IDF) (Ramos, 2003). Third, although we integrate various sources to create a comprehensive aspect categories list, the list may not be universally agreed upon. Future studies may propose other categories that are not included in this essay. Eventually, it calls for a formal representation of a set of aspect categories using techniques such as ontology modeling. Fourth, we enrich the matched CHV terms in reviews to overcome the major limitation of the vocabulary-based approach, which is that the pre-defined vocabularies do not contain relevant terms. However, it is still possible that we do not capture all the relevant terms in our pre-defined vocabularies. Future studies can develop such dictionaries for ABSA analysis. Moreover, we manually group the matched CHV terms into corresponding dictionaries, which is time- and labor-intensive. Future studies can develop programs to automate this grouping process based on similarity values between keywords using models such as Word2vec model (Mikolov et al. 2013).

References

1. Al-Aswadi, F. N., Chan, H. Y., & Gan, K. H. (2020). Automatic ontology construction from text: a review from shallow to deep learning trend. *Artificial Intelligence Review, 53*(6), 3901-3928.
2. Al-Ghuribi, S. M., Noah, S. A. M., & Tiun, S. (2020). Unsupervised semantic approach of aspect-based sentiment analysis for large-scale user reviews. *IEEE Access, 8*, 218592-218613.
3. Al-Mazrou, Y., Al-Shehri, S., & Rao, M. (1990). Principles and practice of primary health care. *General Directorate of Health Centers, 2*(2), 31-42.
4. Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research, 58*(2), 175-191.
5. Ali, F., Kwak, D., Khan, P., El-Sappagh, S., Ali, A., Ullah, S., ... & Kwak, K. S. (2019). Transportation sentiment analysis using word embedding and ontology-based topic modeling. *Knowledge-Based Systems, 174*, 27-42.
6. Arrow, K. J. (1978). Uncertainty and the welfare economics of medical care. In *Uncertainty in Economics* (pp. 345-375). Academic Press.
7. Banerjee, S., Bhattacharyya, S., & Bose, I. (2017). Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business. *Decision Support Systems, 96,* 17-26.
8. Banjar, A., Ahmed, Z., Daud, A., Abbasi, R. A., & Dawood, H. (2021). Aspect-based sentiment analysis for polarity estimation of customer reviews on Twitter. *Computers, Materials & Continua, 67*(2), 2203-2225.
9. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993-1022.
10. Bowling, A., Rowe, G., Lambert, N., Waddington, M., Mahtani, K. R., Kenten, C., ... & Francis, S. A. (2012). The measurement of patients' expectations for health care: a review and psychometric testing of a measure of patients' expectations. *Health Technology Assessment, 16*(30).
11. Brody, S., & Elhadad, N. (2010, June). An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 804-812).
12. Caputo, A., Basile, P., Gemmis, M. D., Lops, P., Semeraro, G., & Rossiello, G. (2017). SABRE: a sentiment aspect-based retrieval engine. *In Information Filtering and Retrieval* (pp. 63-78). Springer, Cham.
13. Chen, M., & Yao, T. (2010, October). Combining dependency parsing with shallow semantic analysis for Chinese opinion-element relation identification. In *2010 4th International Universal Communication Symposium* (pp. 299-305). IEEE.
14. Chen, R., & Xu, W. (2017). The determinants of online customer ratings: a combined domain ontology and topic text analytics approach. *Electronic Commerce Research, 17*(1), 31-50.
15. Chen, Y., & Lee, S. (2021). User-Generated Physician Ratings and Their Effects on Patients' Physician Choices: Evidence from Yelp. Available at SSRN 3796740.
16. Cheng, L. C., & Yang, Y. (2022). The Effect of Online Reviews on Movie Box Office Sales: An Integration of Aspect-Based Sentiment Analysis and Economic Modeling. *Journal of Global Information Management (JGIM), 30*(1), 1-16.

17. Cheng, Y. H., & Ho, H. Y. (2015). Social influence's impact on reader perceptions of online reviews. *Journal of Business Research, 68*(4), 883-887.
18. Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*(3), 345-354.
19. D'Aniello, G., Gaeta, M., & La Rocca, I. (2022). KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis. *Artificial Intelligence Review*, 1-32.
20. Daniulaityte, R., Chen, L., Lamy, F. R., Carlson, R. G., Thirunarayan, K., & Sheth, A. (2016). "When 'bad'is 'good'": identifying personal communication and sentiment in drug-related tweets. *JMIR Public Health and Surveillance, 2*(2), e6327.
21. Daskivich, T. J., Houman, J., Fuller, G., Black, J. T., Kim, H. L., & Spiegel, B. (2018). Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. *Journal of the American Medical Informatics Association, 25*(4), 401-407.
22. Do, H. H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Systems with Applications, 118*, 272-299.
23. Dragoni, M., Federici, M., & Rexha, A. (2019). An unsupervised aspect extraction strategy for monitoring real-time reviews stream. *Information Processing & Management, 56*(3), 1103-1118.
24. Du, J., Xu, J., Song, H. Y., & Tao, C. (2017). Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. *BMC Medical Informatics and Decision Making, 17*(2), 63-70.
25. Eirinaki, M., Pisal, S., & Singh, J. (2012). Feature-based opinion mining and ranking. *Journal of Computer and System Sciences, 78*(4), 1175-1184.
26. Engelseth, P., White, B. E., Mundal, I., Eines, T. F., & Kritchanchai, D. (2021). Systems modelling to support the complex nature of healthcare services. *Health and Technology, 11*(1), 193-209.
27. Fang, X., & Tao, J. (2019, October). A transfer learning based approach for aspect based sentiment analysis. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 478-483). IEEE.
28. Feng, A., Zhang, X., & Song, X. (2022). Unrestricted Attention May Not Be All You Need–Masked Attention Mechanism Focuses Better on Relevant Parts in Aspect-Based Sentiment Analysis. *IEEE Access, 10*, 8518-8528.
29. Ganu, G., Elhadad, N., & Marian, A. (2009). Beyond the stars: improving rating predictions using review text content. In *WebDB* (Vol. 9, pp. 1-6).
30. Gao, G. G., McCullough, J. S., Agarwal, R., & Jha, A. K. (2012). A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. *Journal of Medical Internet Research, 14*(1), e2003.
31. Gao, G., Greenwood, B. N., Agarwal, R., & McCullough, J. S. (2015). Vocal minority and silent majority. *MIS Quarterly, 39*(3), 565-590.
32. Ghose, A., & Ipeirotis, P. G. (2010). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering, 23*(10), 1498-1512.
33. Gray, B. M., Vandergrift, J. L., Gao, G. G., McCullough, J. S., & Lipner, R. S. (2015). Website ratings of physicians and their quality of care. *JAMA Internal Medicine, 175*(2), 291-293.

34. Greaves, F., Laverty, A. A., Cano, D. R., Moilanen, K., Pulman, S., Darzi, A., & Millett, C. (2014). Tweets about hospital quality: a mixed methods study. *BMJ Quality & Safety, 23*(10), 838-846.
35. Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., & Donaldson, L. (2013). Use of sentiment analysis for capturing patient experience from free-text comments posted online. *Journal of Medical Internet Research, 15*(11), e2721.
36. Hedges, L., & Couey, C. (2019). How patients use online reviews. *Software Advice*, Austin, TX, February 11.
37. Hernández-Rubio, M., Cantador, I., & Bellogín, A. (2019). A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews. *User Modeling and User-Adapted Interaction, 29*(2), 381-441.
38. Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 168-177).
39. Islam, M., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. *Health Information Science and Systems, 6*(1), 1-12.
40. Kastrati, Z., Imran, A. S., & Kurti, A. (2020). Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs. *IEEE Access, 8*, 106799-106810.
41. Kordzadeh, N. (2019). Investigating bias in the online physician reviews published on healthcare organizations' websites. *Decision Support Systems, 118*, 70-82.
42. Lin, C., & He, Y. (2009, November). Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management* (pp. 375-384).
43. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*(1), 1-167.
44. Loke, R. E., & Reitter, W. (2021). Aspect Based Sentiment Analysis on Online Review Data to Predict Corporate Reputation. In *DATA* (pp. 343-352).
45. Lu, S. F., & Rui, H. (2018). Can we trust online physician ratings? Evidence from cardiac surgeons in Florida. *Management Science, 64*(6), 2557-2573.
46. Luca, M., & Vats, S. (2013). Digitizing doctor demand: The impact of online reviews on doctor choice. *Cambridge, MA: Harvard Business School.*
47. Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing, 77*(1), 87-103.
48. Luo, L., Duan, S., Shang, S., & Pan, Y. (2021). What makes a helpful online review? Empirical evidence on the effects of review and reviewer characteristics. *Online Information Review, 45*(3), 614-632.
49. McCallum, A. K. (2002). Mallet: A machine learning for language toolkit. *http://mallet.cs.umass.edu.*
50. Metwally, O., Blumberg, S., Ladabaum, U., & Sinha, S. R. (2017). Using social media to characterize public sentiment toward medical interventions commonly used for cancer screening: an observational study. *Journal of Medical Internet Research, 19*(6), e7485.
51. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781.*

52. Moghaddam, S., & Ester, M. (2010, October). Opinion digger: an unsupervised opinion miner from unstructured product reviews. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (pp. 1825-1828).
53. Moghaddam, S., & Ester, M. (2013, May). The flda model for aspect-based opinion mining: Addressing the cold start problem. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 909-918).
54. Mubarok, M. S., Adiwijaya, & Aldhi, M. D. (2017, August). Aspect-based sentiment analysis to review products using Naïve Bayes. In *AIP Conference Proceedings* (Vol. 1867, No. 1, p. 020060). AIP Publishing LLC.
55. Mukherjee, A., & Liu, B. (2012, July). Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (pp. 339-348).
56. Osman, N. A., Noah, S. A. M., & Darwich, M. (2019). Contextual sentiment based recommender system to provide recommendation in the electronic products domain. *International Journal of Machine Learning and Computing, 9*(4), 425-431.
57. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., ... & Eryiğit, G. (2016, January). Semeval-2016 task 5: Aspect based sentiment analysis. In *International Workshop on Semantic Evaluation* (pp. 19-30).
58. Nguyen Thi Ngoc, T., Nguyen Thi Thu, H., & Nguyen, V. A. (2019). Mining aspects of customer's review on the social network. *Journal of Big Data, 6*(1), 1-21.
59. Qiu, G., Liu, B., Bu, J., & Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational Linguistics, 37*(1), 9-27.
60. Rana, T. A., & Cheah, Y. N. (2017, July). Improving aspect extraction using aspect frequency and semantic similarity-based approach for aspect-based sentiment analysis. In *International Conference on Computing and Information Technology* (pp. 317-326). Springer, Cham.
61. Ranard, B. L., Werner, R. M., Antanavicius, T., Schwartz, H. A., Smith, R. J., Meisel, Z. F., ... & Merchant, R. M. (2016). Yelp reviews of hospital care can supplement and inform traditional surveys of the patient experience of care. *Health Affairs, 35*(4), 697-705.
62. Rastegar-Mojarad, M., Ye, Z., Wall, D., Murali, N., & Lin, S. (2015). Collecting and analyzing patient experiences of health care from social media. *JMIR Research Protocols, 4*(3), e3433.
63. Rehurek, R., & Sojka, P. (2011). Gensim–python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic, 3(2), 2.
64. Rouse, W. B., & Serban, N. (2014). Understanding and managing the complexity of healthcare. *MIT Press*.
65. Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In *Proceedings of the First Instructional Conference on Machine Learning* (Vol. 242, No. 1, pp. 29-48).
66. Sahoo, N., Dellarocas, C., & Srinivasan, S. (2018). The impact of online product reviews on product returns. *Information Systems Research, 29*(3), 723-738.
67. Saifee, D. H., Bardhan, I. R., Lahiri, A., & Zheng, Z. (2019). Adherence to clinical guidelines, electronic health record use, and online reviews. *Journal of Management Information Systems, 36*(4), 1071-1104.

68. Saifee, D. H., Zheng, Z., Bardhan, I. R., & Lahiri, A. (2020). Are online reviews of physicians reliable indicators of clinical outcomes? A focus on chronic disease management. *Information Systems Research, 31*(4), 1282-1300.
69. Sarawgi, K., & Pathak, V. (2017). Opinion mining: aspect level sentiment analysis using SentiWordNet and Amazon web services. *International Journal of Computer Applications, 158*(6), 31-36.
70. Schouten, K., & Frasincar, F. (2015). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering, 28*(3), 813-830.
71. Segal, J., Sacopulos, M., Sheets, V., Thurston, I., Brooks, K., & Puccia, R. (2012). Online doctor reviews: do they track surgeon volume, a proxy for quality of care?. *Journal of Medical Internet Research, 14*(2), e2005.
72. Siering, M., Deokar, A. V., & Janze, C. (2018). Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews. *Decision Support Systems, 107*, 52-63.
73. Thet, T. T., Na, J. C., & Khoo, C. S. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of Information Science, 36*(6), 823-848.
74. Tubishat, M., Idris, N., & Abushariah, M. (2021). Explicit aspects extraction in sentiment analysis using optimal rules combination. *Future Generation Computer Systems, 114*, 448-480.
75. Xu, Y., Armony, M., & Ghose, A. (2021). The interplay between online reviews and physician demand: An empirical investigation. *Management Science, 67*(12), 7344-7361.
76. Xue, W., Zhou, W., Li, T., & Wang, Q. (2017, November). MTNA: a neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing* (Volume 2: Short Papers) (pp. 151-156).
77. Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly, 38*(2), 539-560.
78. Zhang, J., & Liu, R. R. (2019). The more the better? Exploring the effects of reviewer social networks on online reviews. *Journal of Marketing Management, 35*(17-18), 1667-1688.
79. Zhao, J., Liu, K., & Xu, L. (2016). Sentiment analysis: mining opinions, sentiments, and emotions. *Computational Linguistics, 42*(3), 595-598.
80. Zhao, N., Gao, H., Wen, X., & Li, H. (2021). Combination of convolutional neural network and gated recurrent unit for aspect-based sentiment analysis. *IEEE Access, 9*, 15561-15569.
81. Zhou, X., Coiera, E., Tsafnat, G., Arachi, D., Ong, M. S., & Dunn, A. G. (2015). Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Studies in Health Technology and Informatics, 216*, 761-765.
82. Zhu, L., Yin, G., & He, W. (2014). Is this opinion leader's review useful? Peripheral cues for online review helpfulness. *Journal of Electronic Commerce Research, 15*(4), 267.
83. Zunic, A., Corcoran, P., & Spasic, I. (2020). Sentiment analysis in health and well-being: systematic review. *JMIR Medical Informatics, 8*(1), e160