NUMERACY, SEVERITY, AND COMMUNICATING RISK: PERCEPTIONS OF PRESCRIPTION PAIN MEDICATION SIDE EFFECTS

By

Jeffrey Cox

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Communication-Doctor of Philosophy

ABSTRACT

NUMERACY, SEVERITY, AND COMMUNICATING RISK: PERCEPTIONS OF PRESCRIPTION PAIN MEDICATION SIDE EFFECTS

By

Jeffrey Cox

This dissertation reports on a study that explored how individuals interpret and quantify verbal descriptions of the risk of side effects from a hypothetical prescription pain medication, as well as what factors affect these interpretations. While the European Union has set out recommendations for how these terms quantifiers (e.g., "very rare," "common") should be interpreted, studies (Cox, 2016; Berry, Knapp, & Raynor, 2002; Knapp, Raynor, & Berry, 2004) indicate that individuals dramatically overestimate these effects' likelihood.

Situated within fuzzy trace theory (Reyna & Brainerd, 1995) the present study assessed how individuals quantify these terms, as well as what internal (e.g., numeracy, existing perceptions of prescription pain medications) and external (e.g., verbal quantifiers used, severity of side effect) factors influence their *gist* and *verbatim* processing of risk information. The study used a between-subjects experimental design: 2 ("common"/"rare") X 2 (adverb/no adverb) X 2 (severity) embedded within an online survey about impressions of prescription pain medications. Findings reveal that individuals' existing, general perceptions of prescription pain medications have a larger impact on their *gist* perceptions than their *verbatim* ones, while their estimates are significantly higher than experts' recommendations. Important differences between the subjective and objective numeracy scales are also found for participants' confidence in their numerical estimates. Other findings related to the study of risk perceptions, as well as implications for practice and policy, are discussed.

ACKNOWLEDGEMENTS

There are many people without whom this dissertation—and I personally—would be much worse off. A dissertation acknowledgement does not do justice to the support and friendship I have received over the years, but perhaps it's a start.

To my parents: You've always been an unwavering source of encouragement, love, and support. Over the years, no matter what I wanted to do with my life, you never tried to convince me to find more *realistic* dreams. Instead, you encouraged me to figure out ways to make them into reality. This is more important to me than you probably know.

To my friends, near and far: I've been lucky to call some wonderful people my friends. You've made all these years of school easier and much more enjoyable ones. I truly treasure the experiences we've had together and I look forward to where we'll continue to go.

To my advisor and doctoral committee members: Your discussions and constructive criticism helped make this study into a much better dissertation, and helped make me a much better scholar. Jim, you've given me support when I wanted it and freedom when I needed it. I very much look forward to continuing to work with you.

To the MSU Communication faculty: Michigan State is the place where I really learned the importance of research and collaborative relationships, and I've met some truly excellent people. Among them: Sandi, thank you for being a voice of reason when I needed it. Frank, I look forward to more informal discussions about movies and books, which helped reconnect me with my favorite non-academic pastimes.

I also owe a huge thanks to the Graduate School at Michigan State University, whose generous Dissertation Completion Fellowship made this research possible.

LIST OF TABLES	v
LIST OF FIGURES	vi
INTRODUCTION	1
THEORY AND LITERATURE	5
Fuzzy trace theory	5
The challenge of numeracy	10
Verbal quantifiers	14
The multiplicative values of adverbs	15
Verbal quantifiers' use in industry and government	17
Preconceptions towards prescription pain medications	18
Severity and estimating risk	21
METHOD	24
METHOD.	
Pre-test.	
Design of main study	
Sample	
Measures	
Analysis	
RESULTS	
Hypothesis 1	
Hypothesis 2	
Hypothesis 3	41
Hypothesis 4	47
Research question 1	
-	
DISCUSSION	49
Verbal quantifiers and participants' numerical estimates	49
Preexisting drug perceptions, and gist and verbatim processing	
Numeracy—objective and subjective measures	55
Severity, common/rare and recall of verbal quantifiers	59
Implications for theory and research	61
Implications for practice and policy	63
Limitations	67
APPENDIX	71
REFERENCES	87

LIST OF TABLES

Table 1: Example questions from the subjective and objective numeracy scales	11
Table 2: Analysis of Variance of the Induction Check Measures for Severity (Pre-test)	26
Table 3: Experimental design of main study	27
Table 4: Correlations between numeracy, confidence, and range	38
Table 5: Analysis of Covariance for Numerical Estimate	42
Table 6: Analysis of Covariance for Perceptions of Side Effect Likelihood	44
Table 7: Analysis of Covariance for Perceptions of Side Effect Worry	46
Table 8: EU Recommendations for Verbal Quantifiers of Side Effect Risk	52

LIST OF FIGURES

Figure 1: Example measure from Cliff (1959)	16
Figure 2: Sample message	27
Figure 3: Quantitative estimate measure	31
Figure 4: High and low quantitative (range) estimate measure	32

INTRODUCTION

Americans' use of prescription drugs has increased dramatically in recent years, with an estimated 59% of US adults reporting having taken at least one prescription drug in 2012, and nearly 15% of US adults taking five or more each month (Kantor et al., 2015). This figure is up from 51% of adults who reported taking prescription drugs in 2000. While these drugs are often necessary to treat serious medical conditions, they come with distinct risks in the form of side effects—secondary, largely undesired reactions. The Food and Drug Administration's (FDA) website states that "all medicines have benefits and risks ... [which] could be less serious things, such as an upset stomach, or more serious things, such as liver damage" (FDA, 2016, pp. 1). The potential danger of these side effects makes it imperative that consumers understand the risks that they face.

Of particular worry in recent years has been a drastic increase in the use, abuse, and death from prescription pain medications. Strong prescription painkillers, such as opioids, are highly addictive and pose a major threat of overdose to users, with more than 1,000 prescription opioid-related hospital visits each day in 2013 (DHHS, 2013). In 1999, prescription painkillers were involved in 30% of drug overdose deaths in the United States, but this number had increased to 60% (or 16,651 deaths) by 2010 (Jones, Mack, & Paulozzi, 2013.) The abuse of these drugs has also taken a financial toll on the American health system, estimated to cost over \$72 billion dollars each year, similar to the amount devoted to treating either asthma or HIV/AIDS (CDC, 2013).

The Centers for Disease Control and Prevention (CDC) claims that the recent increases in prescription drug use can be partly explained by an increase in the marketing of prescription drugs directly to consumers, with "spending on direct-to-consumer advertising for all drugs more

than tripl[ing] between 1996 and 2005, to \$4.2 billion" (CDC, 2015). But how can the way these drugs are described affect individuals' perceptions of their personal risks? More specifically, how do individuals interpret and quantify the risks they find in descriptions of prescription pain medications?

Information about medical risk is sometimes presented to individuals in the form of numbers, such as fractions, percentages, or ratios (Fagerlin & Peters, 2011). Health professionals tend to perceive such numbers as being very precise and easy to understand, so their use is an attempt to communicate information in the clearest and most useful way (Reyna & Adam, 2003; Reyna, 2008). As such, it may seem reasonable that telling an individual that they have a "25.5% risk of prostate cancer" is more useful (for their understanding and future behavior) than simply informing them that they have a "high risk" or an "elevated risk." However, prescription drug advertisements and labeling tend to present information about risk likelihood using verbal descriptions, rather than numerical probabilities. These terms—such as "common" and "very rare"—are often referred to as verbal quantifiers (Fagerlin & Peters, 2011; Hartley, Trueman, & Rodgers, 1984; Newstead & Collis, 1987).

Fuzzy trace theory (Reyna & Brainerd, 1995) posits that individuals make better, healthier decisions with information when they take the bottom-line meaning (the "gist") of the information rather than just the statistical figures (the "verbatim") that they would have to interpret (Fagerlin & Peters, 2011; Reyna & Brainerd, 2008). While a *verbatim trace* is an individual's lasting impression of the presented information (words or numbers) in isolation, a *gist trace* integrates the presented information with prior knowledge and attitudes to form an underlying meaning, such as interpreting a risk as "high" or "low," or "worth the risk."

A better understanding of individuals' interpretations of verbal quantifiers could help risk communicators convey probabilistic risk information in a form that potential users find more meaningful. While no specific guidelines exist in the United States for how to communicate these risks, recommendations have been made in some other countries. For example, the Council for International Organizations of Medical Sciences set out a list of recommendations for how these terms should be used and interpreted when it comes to pharmaceutical product labeling (CIOMS III, 1995). However, a lack of consistency in regulations means that lay people who read the words "rare" and "common" may have quite different interpretations of what these terms mean when compared with the experts who write them.

This dissertation describes a research project that explores how individuals interpret probabilistic terms such as "very common" and "somewhat rare," as used in advertising descriptions of prescription drugs' side effects. Past research (Berry, Knapp, & Raynor, 2002; Knapp, Raynor, & Berry, 2004; Cox, 2016) has shown that individuals tend to dramatically overestimate the likelihood of side effects occurring based on these descriptions, when compared with the European Union recommendations for how these terms are intended to be interpreted (e.g., "very rare" side effects affect .001% to .01% of users) (CIOMS III, 1995). This study aimed not only to assess how individuals interpret these terms, but also what underlying factors influence these interpretations.

To answer these questions, this study employed an experiment imbedded into an online survey programmed in Qualtrics, with a national sample of participants recruited by and directed from Survey Sampling International (SSI). Participants were presented with one of 10 different messages (including a control condition), which varied by the verbal quantifier used, as well as the severity of the listed side effect. Participants then estimated the numerical likelihood of the

listed side effect occurring, before answering other questions about the medication. Other questions assessed participants' numeracy (using two types of measures discussed later), as well as their preconceptions concerning prescription pain medications more generally. This variety of measures was aimed at assessing not only how individuals interpret these verbal expressions of probability, but also how these interpretations are influenced by message factors, and individual traits and predispositions.

The following sections of this dissertation will outline the literature on numeracy, verbal quantifiers, and fuzzy trace theory. Included in this discussion will be the hypotheses derived from these studies' findings. Next, the dissertation will outline the research methods used to explore these questions, before explaining the ways in which the data were analyzed. Finally, a discussion of the implications of this study will be described.

A key aspect of this study's intentions is the understanding that interpretations about one's risk are not a simple stimulus-response from reading a message. Rather, individuals form interpretations within a complex information environment, into which new information is integrated and compared. Individuals who read, watch, or listen to a description of a quantitative risk do not apply these messages' content to their lives in a vacuum, but instead incorporate this new information with their existing knowledge, beliefs, and attitudes about the subject. Therefore, a central aim of this study was to explore these other attributes and their relationships with participants' interpretations of the message.

THEORY AND LITERATURE

Fuzzy trace theory

Fuzzy trace theory (FTT) (Reyna & Brainerd, 1995; Brainerd & Reyna, 1990) examines the interrelationship between memory and reasoning. FTT originates from findings that surprised the theory's founders, demonstrating that reasoning is oftentimes distinct from the accuracy of a person's memory. In particular, the theory was formulated in response to several studies that showed that individuals' memory or recall of specific quantitative facts did not seem to influence their eventual decisions and judgments based on those facts (Brainerd & Reyna, 1990).

Fuzzy trace theory proposes that when people are required to make judgments based on quantitative information, they often do so based on their general impressions (the *gist* representation of the information), rather than focusing on the actual statistics themselves (the *verbatim* representation). This is similar to Daniel Kahneman's (2011) System 1 and System 2 thinking, where some decisions are impulsively and emotionally charged (System 1) while others are the result of careful and logical consideration of the information (System 2). FTT has been applied widely, to better understand phenomena such as false memories and the mental retrieval of information (Reyna, Nelson, Han, & Dieckmann, 2009). Most relevant to the present work, several empirical studies have applied fuzzy trace theory to decision-making, especially in a medical or health risk context (Reyna, 2008; Reyna, 2012).

Fuzzy trace proposes that there are two distinct, but simultaneous cognitive routes through which individuals process decision-based quantitative information. These routes lead to separate representations or *traces* of this information, which inform the decision making process in different ways. *Verbatim traces* are essentially the literal transmission of quantitative

information into one's memory. For example, if someone were told that the side effects of a prescription medication included a 15% chance of experiencing migraines, the verbatim representation would be, "I have a 15% chance of experiencing migraines if I take this medication." Verbatim representations do not include any interpretation of this information, nor any links to existing information in one's memory. They simply lead to a placement of the numbers into one's mind. As such, verbatim traces are thought to relate more to the functioning of one's memory (or retrieval), as opposed to one's reasoning (Reyna & Brainerd, 1995). These traces are also shorter lasting, being susceptible to quick memory decay (Brainerd & Reyna, 1990).

Conversely, *gist traces* are "bottom-line" representations of the information, which draw heavily on an individual's interpretation of new quantitative facts, while integrating them with previously stored information. To continue with the example given above, a gist trace could condense a 15% risk of migraines to be seen as a "likely" or "unlikely" occurrence, depending on factors such as the probability of other side effects occurring, or the prevalence of chronic migraines in the general population. These interpretations are based on integrating the new, quantitative information with what the individual already knows or believes. Additionally, there are often multiple gist traces that are derived from the information. An individual's gist traces in the above example could therefore include additional dimensions, such as "migraines aren't really that painful if you get them," or "I have several friends with chronic migraines, so maybe they're more common than people think." Each of these interpretations, beyond the verbatim understanding of the statistics, can have important implications for how the encoding of information can affect one's decision-making process. Unlike some dual-processing theories,

both of these mechanisms operate simultaneously, as people "extract multiple 'hierarchies' of gist from information" (Reyna, 2008, pp. 3).

One of the central tenets of Fuzzy Trace Theory is that individuals have a preference for relying on gist traces when it comes to making decisions based on probabilities (Reyna & Brainerd, 1995; Reyna, 2008). This theory connects with other perspectives of decision-making, particularly the work of Kahneman & Tversky (1979), upon which Reyna has said fuzzy trace theory was partly based. Take, for example, the famous Asian disease problem that Kahneman & Tversky propose in their research using prospect theory. This mental exercise looks at framing effects for different risk outcomes of a hypothetical disease outbreak. The authors found that individuals who were exposed to a loss frame tended to be more risk seeking (or more likely to take a gamble to avoid a certain loss) than those presented with a gain frame. This ties in with fuzzy trace theory's emphasis on extracting bottom-line information from the data, as people tend to break the different propositions into two general *gist* categories: a situation where some people are sure to die, and another in which there's a chance no one will die. In this case, "because saving some people is better than saving none (a core value), the sure option is preferred" (Reyna, 2008, pp. 855).

Notably, the preference for relying on *gist* representations is not something that just affects inexperienced people. In fact, the preference for *gist* representations increases with individuals' expertise and experience. In a medical decision-making context, physicians have been found to make more accurate and timely diagnoses of patients' ailments when they rely on more general information that has been informed by experience, rather than going through all of the possible "textbook" interpretations of their symptoms (Reyna, 2008).

Fuzzy trace theory can be seen as a reaction to a previously popular perspective on the relationship between memory and reasoning: *memory necessity*. The memory necessity perspective held that memory was a necessary condition for any form of intuitive reasoning. In other words, an individual cannot reason effectively if they don't have an accurate memory about specific and relevant decision-making information. Brainerd and Kingma (1984, 1985) conducted eight experiments that investigated the relationship between memory performance and reasoning performance in school children. Their work demonstrated that, contrary to expectations, a participant's reasoning performance was largely independent of his or her memory performance. In other words, having an accurate and detailed recall of specific quantitative information. The observation that individuals have a hard time making connections between quantitative information and reasoning (Reyna, 2008) has been seen even in highly educated and informed samples, such as physicians (Reyna & Adam, 2003).

One every day example demonstrating the difference between verbatim and gist processing is telling the time with a digital vs. analog clock. A digital clock presents individuals with numbers and requires certain calculations, of a sort, in order to be useful. If one has an appointment at 10:00 a.m. and suddenly discovers that the clock on the wall says 9:45, that individual must first go through several distinct, if fairly quick, steps. First, they must acknowledge the current time and then calculate how much time is left until their meeting (i.e., 15 minutes). Then they must use that information to form an underlying, bottom-line impression ("I need to leave now, so I won't be late.") An analog clock is designed to give individuals an approximate time at a glance, so that they can immediately tell the bottom line (i.e., what the time means in terms of their behavior.) In other words, there are fewer cognitive steps to gaining the bottom-line meaning of an analog (gist) representation of time, rather than a digital (verbatim) one. Additionally, digital clocks often require people to double check more frequently. After all, at first glance, 9:45 and 9:25 look much more similar on a digital clock than they do on an analog one.

At first glance, the theory may appear to be only slightly different from other dualprocessing models, such as the elaboration likelihood model (Petty & Cacioppo, 1986) and the heuristic-systematic model (Chaiken, 1980). There are several factors that make FTT a more relevant theoretical background for this particular study than others. One is that background influences and perceptions are a central part of the theory, posited to have a large impact on individuals' gist processing. This is particularly important for the present study, because background impressions and perceptions of prescription drugs are a large part of the current experiment. Reyna has also stated that numeracy, another central focus of this experiment, is an important factor in gist processing of information (Reyna, 2008). Since risk messages in directto-consumer (DTC) advertisements of pharmaceutical drugs very often use verbal quantifiers to convey quantitative risk, rather than actual numerical figures, this kind of environment would be an excellent environment for the application of this particular theory. Terms such as verbal quantifiers are likely to influence individuals' bottom-line interpretations of the information in the message, contributing to their gist processing of the risk associated with the listed side effect. Additionally, the theory has been used in a number of different contexts, including medical decision-making, and FTT has been described as a "theory of medical decision-making," with a particular relevance for individuals' processing of risk information (Reyna, 2008, pp. 850).

The challenge of numeracy

Numeracy is considered a form of literacy, referring to an individual's ability to understand and apply quantitative information, such as fractions, percentages, decimals, and ratios (Fagerlin & Peters, 2011). Additionally, some researchers have explored the concept of numeracy by expanding this definition to include not only an individual's innate ability when working with numbers, but also his or her preferences for how quantitative information is received and applied (Fagerlin et al., 2007; Zikmund-Fisher et al., 2007). A lack of numeracy is a widespread problem in the United States, with almost 14% of Americans scoring at the "below basic" levels in national assessments of numerical ability (IES, 2003). Notably, low numeracy is a problem that affects individuals with widely varying levels of education and income, and across diverse international populations. Studies have shown that even highly educated professionals such as physicians and lawyers have fairly widespread problems with numerical interpretation (Ghazal, Cokely, & Garcia-Retamaro, 2014). Numeracy is also a problem for national populations with much higher average mathematical skills than Americans, such as in Japan (Okamoto et al., 2012). For the study of communication, this means that numeracy has implications beyond just communicating with specific sub-populations—such as those of lower socioeconomic status or non-native speakers of English. Numeracy is a problem that affects individuals across all social strata.

Numeracy measures fall into two primary categories: objective and subjective scales. Objective measures of numeracy aim to assess a participant's level of numerical ability directly, by asking him or her to make mathematical calculations and conversions (Schwartz et al., 1997; Lipkus, Samsa, & Rimer, 2001). For example, Lipkus, Samsa, and Rimer's (2001) 11-item scale asks participants a series of multiple-choice and open-ended questions that measure their ability to understand mathematical concepts such as fractions, percentages, and converting between the two. However, participants sometimes become frustrated by having to take what is essentially a math test—especially when it's part of an otherwise non-math-related questionnaire—which can lead to lower completion rates or less serious answers to these questions (Fagerlin et al., 2007; Zikmund-Fisher et al., 2007; Fagerlin & Peters, 2011). To address these concerns, Fagerlin, Zikmund-Fisher, and colleagues (Fagerlin et al., 2007; Zikmund-Fisher et al., 2007) developed the subjective numeracy scale (SNS), which seeks to address numerical ability differently than direct assessments of skill. In particular, the SNS asks participants to self-report their ability to understand and use a variety of mathematical concepts, as well as their preferences for receiving and using quantitative information. The 8-item SNS has been found to have both a high internal reliability (a=0.75) and a fairly strong correlation with Lipkus, Samsa, and Rimer's (2001) objective scale (r=0.53). A table that gives examples for how each scale measures ability with the mathematical concepts of fractions/proportions and percentages is show in Table 1.

Mathematical concept	Example from objective	Example from subjective
	numeracy scale	numeracy scale
Fractions/proportions	If Person A's chance of getting a disease is 1 in 100 in ten years, and Person B's risk is double that of A's, what is B's risk?	How good are you at working with fractions?
Percentages	If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1,000?	When you hear a weather forecast, do you prefer predictions using percentages (e.g., "there will be a 20% chance of rain today") or predictions using words (e.g., "there is a small chance of rain today")?

Table 1: Example questions from the subjective and objective numeracy scales

In an unpublished study, Cox (2016) found notable differences in the predictive power of the objective and subjective numeracy scales. While higher scores on the subjective scale were correlated with a higher certainty in participants' estimates (i.e., a narrower difference between a their high and low estimates), response certainty appeared to be associated with *lower* scores on the objective scale. In other words, the higher a participant scored on the objective numeracy measure, the more likely they would have a wide range between their high and low numerical estimates (or less certainty in their responses). These findings appear to indicate that objective and subjective numeracy scales—two constructs that are sometimes discussed as being interchangeable measures—may not actually be measuring the exact same things. At first glance, objective numeracy is more of a direct evaluation of one's math ability, while subjective measures may appear more similar to mathematical self-efficacy or confidence. Cox's (2016) study could indicate that the two scales may have differing relationships with various outcome measures, including individuals' confidence or certainty in their numerical estimations of probability.

Several studies have demonstrated that individuals who lack basic numerical skills have difficulty understanding quantitative information and making real-world applications (Fagerlin & Peters, 2011; Reyna, Nelson, Han, & Dieckmann, 2009; Nelson, Hesse, & Croyle, 2009). When those with low numeracy are presented with quantitative information, they sometimes tend to ignore or discount it, instead basing their impressions on other aspects of the message (such as visualizations or explanatory text) (Fagerlin & Peters, 2011; Reyna & Brainerd, 2008). This aversion to quantitative information can lead to unhealthy behavior and has been associated with issues such as low medication compliance and a hesitance to seek necessary medical treatment (Reyna, Nelson, Han, & Dieckmann, 2009).

Low numeracy is a serious barrier to effectively communicating quantitative and probabilistic information to the public more generally, but this problem is accentuated in situations involving health or risk information. Health risks often require that individuals have a keen understanding of their susceptibility to negative outcomes. However, widespread unease with numbers makes it difficult for risk communicators to clearly convey important likelihoods. One of the ways that some researchers have proposed addressing this problem is by replacing numerical likelihoods, such as probabilities or percentages, with descriptive words or phrases referred to as verbal quantifiers (Reyna, 2008).

This literature leads to the first set of hypotheses of the study:

H1a: Participants with higher subjective and objective numeracy scores will report more confidence in their numerical estimates. This is due to the literature that suggests that individuals who score higher on numeracy measures tend to have better mathematical abilities, as well as more comfort and familiarity with using numbers.

H1b: Participants with higher subjective and objective numeracy scores will have a narrower range between their high and low numerical risk estimates. This is because individuals who have higher mathematical ability may have greater certainty about their estimates, leading them to put a more precise range of possible interpretations.

H1c: Subjective numeracy will be a better predictor of estimation confidence than objective numeracy. This is drawn from the different natures of the objective and subjective numeracy scales, which seemed to be reinforced by Cox's (2016) study. In that experiment, each of these scales appeared to be predictive of different things, suggesting that they were not measuring the exact same construct. The subjective numeracy scale asks for respondents' self-

reports about their comfort and preferences for using numbers. Therefore, this scale should be a better predictor of individuals' confidence in their numerical estimates.

Verbal quantifiers

Verbal quantifiers, sometimes known as subjective or linguistic probabilities, are words that are used to represent the likelihood of an event occurring (e.g., "often," "sometimes," "rare," "common"). This is distinct from verbal conveyances of exact numbers, such as "a hundred," because their exact value is open to interpretation. Verbal quantifiers are used for a number of reasons, including ease of information coding, a lack of precise numerical data for reference, or in an attempt to communicate with those who lack adequate numerical skills (Wallsten, Fillenbaum, & Cox, 1986; Fagerlin & Peters, 2011). One of the concepts of fuzzy trace theory is that individuals may find verbal expressions of probability more useful than numerical ones, because they can convey a more direct underlying meaning (e.g., "this risk is high") (Reyna, 2008).

While verbal quantifiers exhibit some clear advantages, they tend to be interpreted quite differently by different individuals (Hartley, Trueman, & Rodgers, 1984; Wallsten, Fillenbaum, & Cox, 1986). Interpretations are often inconsistent and it can be hard for individuals to precisely quantify these terms. In Cox's (2016) unpublished study, the range of interpretations was wide for each of six verbal quantifiers (e.g., "very rare," "somewhat common"), with standard deviations reaching high levels in some cases (e.g., SD=311.630 on a 1,000-point sliding scale). While the order of interpretations (e.g., "rare" having a higher likelihood than "very rare") was fairly consistent, this supports evidence that individuals perceive these words

very differently. The consistency of order is important, however, as it indicates that participants believe these terms to have differing values.

This literature leads to the study's second hypothesis:

H2: Participants' numerical estimations of risk will increase as the description uses larger verbal quantifiers (starting with "very rare" as the least common, to "very common" as the most common.) For example, it is expected that participants will give lower estimates for a side effect that is described in a message as "very rare," versus the same kind of side effect that is described as simply "rare."

The multiplicative values of adverbs

In his article *Adverbs as Multipliers*, Cliff (1959) examined the multiplicative effects that adverbs had on adjectives' elicitation of participants' positive or negative valence towards others. Unlike the current study, which uses these word pairs in combinations that may affect a participant's perceptions of probabilistic risk, Cliff used adjectives (e.g., *evil, bad, pleasant, charming*) and adverbs (e.g., *slightly, rather, quite, extremely*) to measure attitude perceptions towards another individual. The main focus of Cliff's work was whether or not adverbs could be isolated for their multiplicative values across adjectives. Put more simply by Cliff himself in a later article, "various adjectives occupy different positions on a continuum, and various adverbs have different 'multiplying values,' so that the function of the adverb is to move adjectives up or down the continuum in a regular way" (Cliff, 1972, p. 176).

Cliff conducted several experiments asking college students to rate adverb-adjective word pairs—such as "very respectable," "somewhat agreeable," or "slightly charming"—in terms of their positive or negative valence, along a -5-to-+5 scale. An example is shown in Figure 1.

Results suggested that some adverbs could be isolated for their multiplicative values. For example, the adjective *good* was found to have a standardized valence of 1.078 for students at Wayne State University, 1.158 at Princeton University, and 1.075 at Dartmouth University. The adverb *slightly* was found to have a multiplicative value of .555, .538, and .559 for students at these schools, respectively. He also notes that, "the fit was excellent for the adverb matrices, but that the adjectives, while good, [were] less noticeably exact" (Cliff, 1959, p. 38).

Figure 1: Example measure from Cliff (1959)



Cliff's findings are significant because they demonstrate that, at least under certain circumstances, the use of certain word combinations can lead to some consistency in how individuals interpret these terms. If risk communication practitioners and researchers can test this experimentally in a health risk context, it could have important implications for how risk and health information is presented to the public. In Cox's (2016) preliminary study, word pairs of verbal quantifiers ("very common," "very rare") were analyzed to see if the adverbs could be isolated for their multiplicative power. In contrast with Cliff's predictions (and findings), consistencies in participants' interpretations of these adverb terms were not found. Regardless, individuals perceived each of these word pairs as have consistently differing values, so further study of participants differentiation between these words (e.g., between "rare" and "very rare") is warranted.

Verbal quantifiers' use in industry and government

The use of verbal quantifiers is widespread in the pharmaceutical and medical fields as a way of communicating probabilistic information to individual patients or the general public (FDA, 2017). For example, a content analysis by Kaphingst et al. (2010) found that 57% of studied DTC ads used qualitative terms (e.g., "rare") to describe side effect frequency, while only 4% described side effect risks in quantitative terms (e.g., "one in ten"). The FDA has rules on what aspects of prescription drugs can and must be included in advertisements directed at consumers. For example, if the advertisement includes information on the drug's function and efficacy, it must also include information on possible risks associated with taking the drug, including risks from side effects (FDA, 2016.) Failing to do this can result in a "warning letter" being sent from the FDA to pharmaceutical companies whose advertisements are not in compliance with the rules. However, the use and endorsement of verbal quantifiers in professional medicine and government communications is inconsistent. In several US government reports, these terms are dismissed as inaccurate and ill advised. A 2006 US Department of Health and Human Services guidance-for-industry report on pharmaceutical labeling claimed, "the terms 'rare,' 'infrequent,' and 'frequent' do not provide meaningful information about the frequency of occurrence of adverse reactions," warning against terms "for which there are no commonly understood parameters" (USDHHS, 2006, p. 9).

Although there is a lack of consistency in their use, and the basis upon which they were created is vague, some government regulatory agencies outside of the US have recommended that particular word combinations be used to represent specific quantitative risks. For example, in the European Union, the Council for International Organizations of Medical Sciences' (CIOMS) III report recommended the following guide for listing drug side effects: "very

frequent or common" (>10%); "frequent or common" (1 to 10%); "infrequent or uncommon" (.1 to 1%); "rare" (.01 to .1%); "very rare" (.001 to .01%); "exceedingly rare" (<.001%) (CIOMS III, 1995). These recommendations are echoed in the European Commission's 2009 pharmaceutical Summary of Product Characteristics (SmPC) report (European Commission, 2009).

Preconceptions towards prescription pain medications

Cox's (2016) study included several thought-listing questions (Cacioppo, Hippel, & Ernst, 1997) that asked participants to write their immediate impressions after reading prescription drug descriptions. A relatively simple reading of these responses reveals two primary themes that could be seen to influence participants' perceptions of the drug's risk. Participants expressed fear or concern over the drug's potential for addiction ("is there a possibility of addiction?"; "I don't want to become dependent on it") and feelings of distrust or anxiety about drugs more generally ("I don't like drugs, prescription or not;" "a new expensive drug for others to make money").

Concerns such as these are supported by recent data. According to the American Society of Addiction Medicine, "of the 21.5 million Americans 12 or older that had a substance use disorder in 2014, 1.9 million had a substance use disorder involving prescription pain relievers" (ASAD, 2016, pp. 1). While the description given for this prescription pain reliever does not specify it as an opioid, several respondents from Cox's (2016) study presumed as much. "Is it an opiate or not?" one asks. "If it's an opiate it will be additive," another adds. This indicates not only that a number of individuals have more detailed knowledge of prescription painkillers than might be expected, but that this knowledge contributes to their aversion to these kinds of drugs.

The fuzzy trace theory assertion that individuals tend to rely on gist traces, and that such traces are highly influenced by previous beliefs and attitudes, means that participants' responses to these questions could have important relationships with how they interpret the risk information they read in the messages in this study. Individuals' beliefs about negative aspects of prescription pain medications more generally—such as their overall danger and the likelihood of users becoming addicted to them—most likely would contribute to their existing mental approach (or schema) to information regarding prescription pain medications' risks. As such, these baseline beliefs may play a significant role in their processing of new risk information. Because of this, it should be important to include these variables into any analysis that investigates the main factors' (i.e., specific verbal quantifier and severity manipulations) effects on risk perceptions. Therefore, they will be included as covariates in this study's analyses of factors influencing risk interpretations from the listed side effects of the described drug. This proposed relationship leads to two hypotheses:

H3a: Participants who feel that prescription painkillers are dangerous (as reported through Likert-style questions) will have greater perceptions of risk of side effects from the described medication than those who feel they are less dangerous.

H3b: Participants who feel that prescription painkillers pose a significant risk for addiction will perceive greater risk of side effects from the described medication than those who feel they pose a smaller risk.

Importantly, individuals' baseline interpretations of the overall risks of prescription pain medications are not formed entirely on their own, but also by receiving information from other sources. Therefore, it is crucial to be able to take stock of not only participants' existing beliefs about prescription medications, but the sources of information. This way, we may be able to

have a better understanding of where people are learning about prescription pain medication use and abuse, and the effect that these different information sources may have on their baseline beliefs about the drugs. One major place (or set of places) where people turn for information is the news media. In particular, news stories about prescription painkiller abuse and addiction has been prominent in recent years. Many news sources, such as the Washington Post and New York Times, have referred to the escalating problem of prescription painkiller addiction in the United States using dramatic terms such as "epidemic" (NYT, 2017). It is likely that greater exposure to news coverage of pain medication abuse would have an effect on individuals' beliefs about the drugs more widely. In turn, this could have a significant impact on their interpretations of the risk of side effects from the messages they read in this study. This leads to the following hypothesis:

H3c: Participants who report having seen frequent news coverage about prescription pain medication addiction and abuse will have greater perceptions of the risk of side effects from the described medication than those who report a smaller frequency of news viewership on this issue.

Another source of information about prescription pain medications that is likely important is a participants' friends and family. Since pain medication addiction and abuse have been a widespread problem throughout the US, many individuals likely personally know or know of people who have suffered from this problem. Certainly, many individuals are likely to have spoken with friends and family about the issue, whether because people they personally know are affected, or because of the large amount of news coverage that has been devoted to the issue in recent years. Therefore, the frequency of personal interactions with friends and family about

this issue could be an important contributor to their baseline beliefs about prescription pain medications. This leads to the following hypothesis:

H3d: Participants who report having spoken more frequently with friends and family about prescription pain medication addiction and abuse will have greater perceptions of the risk of side effects from the described medication than those who report less frequent interactions about this topic.

Severity and estimating risk

One factor that is believed to influence individuals' risk estimates is their perception of the severity of the listed side effect. Studies have shown that participants pay particularly close attention to a side effect's severity, and that this can influence their judgments based on these risks (Berry et al., 2004 ; Wallsten, Fillenbaum, & Cox, 1986). Furthermore, severity is an important factor in an important health communication model, the Theory of Planned Behavior (Ajzen, 1985). In both of these frameworks, severity plays a key role in determining an individual's health behavior. Although clearly important, the manner in which severity affects perceptions of risk likelihood is up to some debate, with studies alternatively showing that likelihood estimates could potentially be adjusted upwards or downwards for conditions considered to be more severe.

Bonnefon & Villejoubert (2006) conducted an experiment that explored how individuals interpreted a physician's descriptions of their likelihood of having either a low- or high-severity medical condition. The experiment tested how these severity conditions affected participants' interpretations of the term *possible*. Participants were asked to imagine their family physician had told them they would *possibly* develop either insomnia or deafness in the following year.

These two conditions were chosen because of their similar prevalence in the target population (i.e., in France, deafness and insomnia were found to have national incidence rates around 4%). Findings showed that participants tended to overestimate the likelihood of having the severe condition (deafness) more than ones that were widely considered less severe (insomnia). While participants' perceptions could have been due to their beliefs about which ailment was more common in the general population, the authors interpreted this to be due to patients' interpretation that such language was physicians' way of providing a "hedge," or way to "safeguard the feelings of people who are receiving face-threatening news" (Bonnefon & Villejoubert, 2006, pp. 750). However, as this study was done in a face-to-face situation regarding medical conditions, it is quite a different context than the one this dissertation explores. While this study is certainly relevant to the current research, other, perhaps more relevant studies indicate that individuals may have lower estimations for what they perceive to be low-severity side effects.

A series of studies conducted by Dianne Berry and colleagues (Berry, Knapp, & Raynor, 2002; Knapp, Raynor, & Berry, 2004) have looked directly at how severity influences individuals' interpretations of risk from drugs' side effects. For example, Knapp, Raynor, and Berry (2004) conducted an experiment where participants estimated the likelihood they would be affected by either high- or low-severity side effects, where descriptions either used numbers (e.g., "this side effect occurs in 0.04% of people who take this medicine) or verbal descriptions (e.g., "this is a rare side effect"). Participants consistently reported lower likelihoods for more severe side effects, but there was an important confound. In this study, severe side effects were always described as "rare," while mild side effects were always described as "common." While this make it hard to separate the effects of the severity from the description, this does tend to

reflect the real incidence of prescription drug side effects. In other words, more severe side effects almost always are more rare than less severe ones. These conflicting findings lead to the following hypotheses and research question.

H4a: Participants in the high-severity conditions will have better recall of the exact verbal quantifier than those in the low-severity conditions. This is because higher severity side effects will likely worry participants more, prompting them to pay more attention to the description and remember them better.

H4b: Participants in the *rare* conditions will have a better recall of the exact verbal quantifier than those in the *common* conditions. This is largely based on the results of a preliminary study (Cox, 2016) which found that participants seemed to pay more attention to the messages in the *rare* conditions, including having a much greater chance of remembering the exact verbal quantifier than in the common or control conditions.

RQ1: Will participants report higher numerical estimates for *high severity* or *low severity* side effects?

METHOD

Pre-test

A pre-test of the main study was primarily conducted as an induction check, so that it could be clear that participants interpreted the two severity conditions as being significantly different from one another. The side effects of intense nausea and vomiting (high severity) and dry mouth (low severity) were chosen because it was thought that participants would consider these two to be of drastically different severities. This was based on past studies of pain medication users, who had been asked to rate which side effects they felt were the most severe (Palos et al., 2004). If participants did not perceive a large difference between these conditions, this would be a significant shortcoming of the study. In other words, little could be meaningfully said about the effect of severity in an experiment where the two severity conditions were considered to be too similar to each other for participants to see them differently. Therefore, it was decided that a separate pre-test was needed in order to make sure that participants "took" to this induction, before the main study was rolled out.

The pre-test sample consisted of a total of 121 participants taken from Survey Sampling International (SSI) a nationally representative online sampling firm. Of these participants, 44.6% were female and 55.4% were male. The age ranges were fairly consistent between 25-29 years old (14.5%), 30-34 years old (16.5%), 35-39 years old (18.2%), 40-44 years old (17.4%), 45-49 years old (15.7%), and 50-55 years old (17.4%). 62.8% had less than a college education, 21.5% had a bachelor's degree, and 13.2% had a post-graduate degree. 70.2% of participants identified as white/Caucasian, 13.2% as African-American, and 14% as Hispanic. 73.9% were employed for pay at the time of the study. Since the main purpose of the pre-test was to check the induction, and not to test any of the hypotheses related to other aspects of the message description (i.e., the verbal quantifier), only two message conditions were included. Neither of these included a verbal quantifier, with messages stating simply "side effects include (dry mouth/intense nausea and vomiting)." Four separate, nine-point semantic differential questions were used to assess participants' perceptions of the severity of the side effects mentioned in the messages they read. Participants were asked to rate the severity of the listed side effect along the following four scales: not severe—severe, not serious—serious, not upsetting—upsetting, and mild—not mild.

Analyses of variance (ANOVA) were used to analyze the difference between participant's mean responses to each of these items, based on their severity condition. This ANOVA is shown in Table 2. For the severe—not severe item, the mean for the low severity condition was 2.34 (SD=1.56), while it was 6.15 (SD=2.6) for the high severity condition. An ANOVA showed that $F(_{1,119})=95.59$, p < .001. The low severity and high severity means for the not serious—serious item were 2.51 (SD=1.64) and 5.83 (SD=2.55), respectively. The ANOVA for this item was $F(_{1,118})=72.40$, p < .001. For the not upsetting—upsetting item, the means were 2.46 (SD=1.58) for low severity and 6.32 (SD=2.53) for high severity. The ANOVA for this item was $F(_{1,118})=101.56$, p < .001. The mild—not mild item had low severity and high severity means of 2.93 (SD=2.11) and 5.74 (SD=2.66), respectively. The ANOVA for this item showed that $F(_{1,117})=41.10$, p < .001. These analyses reveal that there were highly statistically significant differences for each of the four induction check measures for severity. Therefore, the induction check demonstrated that participants clearly perceived a large difference in the severity of the dry mouth and intense nausea and vomiting messages.

		Sum of				
		Squares	df	Mean Square	F	Sig.
Not severeSevere	Between Groups	438.100	1	438.100	95.585	.000
	Within Groups	545.420	119	4.583		
	Total	983.521	120			
Not seriousSerious	Between Groups	331.041	1	331.041	72.399	.000
	Within Groups	539.551	118	4.572		
	Total	870.592	119			
Not upsettingUpsetting	Between Groups	447.563	1	447.563	101.557	.000
	Within Groups	520.029	118	4.407		
	Total	967.592	119			
MildNot mild	Between Groups	234.251	1	234.251	41.099	.000
	Within Groups	666.858	117	5.700		
	Total	901.109	118			

 Table 2: Analysis of Variance of the Induction Check Measures for Severity (Pre-test)

Design of main study

This online study was conducted as a 2 (adjectives "rare" vs. "common") X 2 ("very," vs. no adverb) X 2 (high severity vs. low severity) full factorial between-subjects experimental design, with the addition of two control conditions (for *high severity* and *low severity* conditions) that said simply "side effects include ______." This control is different from the control condition used in Cox (2016) (i.e., "possible side effects include...") for two reasons. One is that the term "possible" had been used in an attempt to find a numerically value-neutral word, but participants were found to perceive it as a midpoint between "somewhat rare" and "somewhat common." Since participants actually showed consistency in assigning it a numerical value, it was deemed unsuitable as a true control. Additionally, given the new inclusion of questions asking for participants' general predisposition towards pharmaceuticals and apprehension toward untested drugs, having a control with no indication of likelihood could give a view into these individuals' baseline interpretations of pharmaceutical drugs' side effect likelihood.

Participants were presented with a message that gave a brief description of the pain-relief drug, and then listed one of two side effects (one high severity, one low severity) with a varied verbal description of its likelihood of affecting users. As this was a between-subjects design, each participant was only exposed to one message condition, so that they would not adjust their interpretations in a relative sense, by comparing the messages. An example message is shown in Figure 2. A diagram of the experimental design for the main study is shown in Table 3.

Table 3: Experimental design of main study

HIGH SEVERITY	"Very"	No adverb
Common	Very common	Common
Rare	Very rare	Rare

Control: Side effects include intense nausea and vomiting.

LOW SEVERITY	"Very"	No adverb
Common	Very common	Common
Rare	Very rare	Rare

Control: Side effects include dry mouth.

Figure 2: Sample message

Manufacturer's product description:

Prescription Soulagis caplets provide temporary relief of moderate-to-severe pain. A single caplet provides up to 12 hours of pain relief. Soulagis is intended for adults and children 12 years and older. Take with a full glass of water.

Rare side effects of this medication include intense nausea and vomiting.

There are several reasons why drug side effects are being used here. Paul Slovic's (1987)

work on risk perception indicates that people often perceive more risk from dangers they see as

being out of their control. Side effects, of one kind or another, are commonly present in medications. This risk-taking may been seen as different from smoking cigarettes, for example, because oftentimes drugs with unpleasant side effects are necessary to maintain health and therefore cannot simply be avoided. This may lead people to believe that side effects are dangers that are relatively out of their hands, especially if it's a medication that they perceive as being important. Furthermore, most individuals have had some experience using pain relievers (whether prescription or over-the-counter), so it is a situation that is relevant to the general public. Perhaps more importantly, advertisements for prescription drugs very often include risk information in probabilistic terms, such as *rarely reported* or *commonly reported*, so this is a realistic topic with which to test their use (Avorn & Shrank, 2009; Berry, Raynor, & Knapp, 2003).

Sample

A total of 769 participants were recruited for this study, but filter questions at the beginning of the survey left a final usable sample of 712 participants. There were a total of three filter questions. The first asked if participants understood the information provided in the consent form and agreed to be a part of the study. The second filter question asked if they were between the ages of 25 and 55. This was an important factor in looking at responses to prescription medication use. Since prescription use often increases as people get older, a student sample would likely give a non-representative view of the public's perceptions. The final filter question asked participants to pledge to "carefully read all of the information in this survey and provide thoughtful and honest answers to all of the questions. Another question asked if

participants had taken a pain medication before (either prescription or over-the-counter), but they were not filtered out based on their response.

The final sample was roughly equal between the number of men (45.6%) and women (54.1%). Ages were also roughly equivalent between those aged between 25-29 years old (16.5%), 30-34 (17.4%), 35-39 years old (17.4%), 40-44 years old (16.2%), 45-49 years old (15.8%), and 50-55 years old (16.7%). In terms of education, 55.9% of participants had less than a college degree, mirroring the 65% of the US population with less than a collegiate education (Ryan & Bauman, 2016). 29.4% of participants had a bachelor's degree, while 12.9% had a graduate degree. 77.9% of participants identified as Caucasian, 10.7% as African-American, 8.9% as Hispanic, 1.1% as multiracial, and 0.6% as Native American. 64% of participants were presently employed for pay.

This experiment was conducted as an online survey through the survey administration tool Qualtrics. Participants were recruited from on online participant recruitment firm, Survey Sampling International (SSI). This was largely because it has been successfully used in many kinds of recent studies and allows specific customization of participant demographics, including a more representative and generalizable sample than a student one would allow. They also recruit demographically diverse participants from around the nation to take surveys and allow researchers to customize their samples based on a number of demographics. All participants recruited were between the ages of 25-55. One of the advantages of using SSI is the ability to stop sample collection at different times throughout the process, in order to check demographics. For example, if a disproportionate number of participants are too highly educated, or there are fewer Hispanics than is nationally representative, the participants are filtered to make sure that demographics that are under-represented in the sample are boosted for the rest of the collection.

Therefore, close attention was paid to having the sample reflect demographic census data for the United States population, especially in terms of participants' gender, ethnicity and educational level.

Measures

Since a major aspect of this study was to experimentally explore the differences between participants' *verbatim* interpretations of risk messages and their *gist* interpretations, several different kinds of measures were developed to assess participants' risk perceptions. One set of questions aimed to gauge participants' *verbatim* interpretations by asking them to give quantitative estimates of what proportion of individuals who took the drug in question would be affected by the listed side effect, based on the drug description they had been presented. Another set of questions aimed to assess individuals' bottom-line, or *gist* risk perceptions. This allowed for an examination of how a number of different variables related to participants' *verbatim* and *gist* interpretations of risk.

Two measures were used to assess participants' *verbatim* quantitative estimates. The first asked them to report how many possible users of the medication, out of 1,000, they believed would be affected by the side effect described in the message. To assess this, participants were presented with a sliding scale that ranged from 0 to 1,000 and were asked to move the slider along the scale to indicate what proportion of users they believed would be affected. Next to the scale was a number that corresponded to the point of the slider. Note that the slider was not present before participants clicked on the scale. This was done so that participants would not simply move the slider up or down from, for example, a midpoint. The sliding scale measure was employed because some research has shown that people often find it easier to understand
relative frequencies (e.g., 1 in 1,000), rather than percentages (e.g., 0.1%) (Gigerenzer et al.,

2008). A sliding scale can offer participants a greater level of nuance than choosing one from a number of preset options. An example of the slider measure that participants were asked to use is shown in Figure 3.

Figure 3: Quantitative estimate measure

If 1,000 patients took Soulagis (the prescription you just read about), around **how many** of these patients would you expect to experience the listed side effect? (Please indicate the number using the slider below).

0	100	200	300	400	500	600	700	800	900	1000
Users affected by the side effect										

Since probabilities are often thought of in terms of a range of possible outcomes (e.g., if a fair coin is flipped 100 times, it will likely come up heads between 40 and 60 times), a second question asked participants to use two sliding scales to give estimates of what they believed the high and low possibilities would be. This was primarily done in order to assess the range of interpretations that participants had. Having participants' range of interpretations was desirable because it would allow for a better understanding of the precision or certainty of their estimates. As noted earlier, Cox (2016) used the difference (or range) between participants' high and low numerical estimates as a proxy for estimate confidence. This dissertation also included a more direct measure of estimate confidence, with a self-report question asking participants to indicate their level of confidence in the estimate they have given. This measure also appears to be more consistent with the subjective numeracy scale, since the SNS measures individuals' self-reported comfort with using numbers.

For example, an individual who had only a 25-point difference (e.g., between 100 and 125 out of 1,000 people would be affected) between their high and low estimates could be said to be more certain of their estimate than an individual who had a 250-point difference (e.g., between 100 and 350 out of 1,000 people would be affected). In addition to an analysis of their range of estimates, participants were asked a question about their level of confidence in their estimate after each of the quantification questions. In effect, this allowed for two different ways of measuring participants' level of confidence or certainty in their risk likelihood estimate. A depiction of the range-estimate measure is shown in Figure 4.

Figure 4: High and low quantitative (range) estimate measure

Sometimes we are uncertain about how many times something will occur, so we indicate a **range** of likely values. For example, if you flipped a coin 100 times, you might predict that it is likely to come up "heads" somewhere between 40 and 60 times.

Similarly, in the question below, please indicate the **range** of users (out of 1,000) who are likely to experience the side effect after taking this prescription.

"If 1,0 users	1,000 people took this prescription, I would expect between _ ers to experience the listed side effect."							and	of these	
0	100	200	300	400	500	600	700	800	900	1000
Low	estimate									
High	estimate	_								

In order to assess participants' bottom-line or *gist* interpretations of side effect risk from the message they read, two Likert-style questions aimed to gauge their risk perceptions in a more general sense. The first of these was "If I took this drug, I would be likely to experience side effects." This was included because it is asking for a more general interpretation of essentially the same construct that the quantification questions were asking—participants' perceptions of the likelihood of an individual being affected by the side effects. However, instead of asking them to report a specific estimate (or set of estimates), participants could report in more general terms the overall likelihood they estimated.

The second *gist* perception question was "If I took this drug, I would be worried about side effects." This was considered to be an even more general question for assessing participants risk perceptions of the drug, since it is asking more abstractly about their level of *worry* about the medication's side effects. However, worry about a drug's side effects is a wide construct that would likely not only capture participants' perceptions of the *likelihood* of being affected by the medication's side effects, but also perceptions of *severity* from the side effect, and perhaps other factors. In fact, studies have indicated that individuals' level of worry about a given risk is more driven by their perceptions of how severe the outcome could be (regardless of its likelihood of happening) (Loewenstein et al., 2001).

Since numeracy is considered to be an important factor influencing individuals' risk assessments—in both their *gist* and *verbatim* processing—a measurement of participants' numerical ability was key to this study (Reyna, 2008). As mentioned in the review of relevant literature, there are several different ways of assessing an individual's numeracy, each with its own strengths. While objective numeracy measures can offer a clear look at one's mathematical ability by asking them to complete math questions in real time, they are time-consuming and often aggravating for people to complete (Fagerlin et al., 2007). The open-ended-response nature of many of the questions also allows participants to write answers that are not the result of any actual mathematical thought.¹ The specific scale that was used in this study was Lipkus,

¹ Indeed, this was found to be the case for participants in the present study. Since several of the objective numeracy measures required participants to fill in answers to open-ended mathematical questions, a number of participants appeared to bypass actually completing the questions by writing answers such as "I don't know," or "I hate math."

Samsa, & Rimer's (2001) 11-question scale. Of these, three questions are multiple choice, while eight are open-ended questions. Examples of questions in this scale include, "If the chance of getting a disease is 10%, how many people would be expected to get this disease out of 1,000?" and "If Person A's chance of getting a disease is 1 in 100 in ten years, and Person B's risk is double that of A's, what is B's risk?"

The subjective numeracy scale here is the 8-item scale developed by Fagerlin et al. (2007). Each of these questions asks participants to report their responses to certain statements about mathematic ability or preferences, along a six-point scale. Examples of these types of questions included, "When you hear a weather forecast, do you prefer predictions using percentages (e.g., "there will be a 20% chance of rain today") or predictions using words (e.g., "there is a small chance of rain today"?); and "how good are you at figuring how much a shirt will cost if it is 25% off?" After reverse coding one of these items, they were averaged into a single scale.

A preliminary study by Cox (2016) also found intriguing differences in the predictive power of each of these measures, suggesting that, while they may not be measuring the exact same construct, they are both highly useful measures. This finding, in addition to the strengths of each type of scale individually, led to the inclusion of both types of measurement in this study. This way, the differences between subjective and objective numeracy measures (and perhaps different individual difference variables) could be measured and compared.

Since *gist* processing of risk information relies on the concept that individuals' interpretation of information is incorporated into and influenced by existing beliefs, opinions and experiences, several other questions aimed to assess participants' baseline attitudes and beliefs

Other answers to these questions included colorful responses such as, "I'm not taking your math test, buddy," and—the author's personal favorite—"whomever (sic) thought of this survey needs a life."

about prescription painkillers. Many of these questions were borrowed from a survey conducted jointly by the Boston Globe and Harvard School of Public Health in May 2015, which focused on prescription painkiller abuse and addiction (BG-HSPH, 2015). Other questions were developed with the aim of assessing as many of the factors as possible that could potentially influence participants' risk perceptions.

Two questions gauged participants' beliefs about prescription painkillers more generally. The first question asked for their level of agreement with the statement "Prescription painkillers are dangerous to those who take them." A second question asked participants to respond to a question asking, "How likely do you think it is that a person taking a strong prescription pain medication will become addicted to it." Both of these questions could be important covariates in understanding participants' existing impressions about the dangers of using prescription painkillers—potentially important influences on their risk perceptions of a message about a new prescription pain medication. Two additional questions aimed at assessing where participants might be getting information about prescription pain medication addiction and abuse. These asked how frequently participants had seen news coverage on the issue, and how often they had spoken with friends and family about painkiller abuse.

Analysis

Hypothesis 1 was analyzed with a Pearson correlation to examine the relationship between each of the two numeracy measures, and participants' confidence/certainty in their estimates. Hypothesis 2 was analyzed with an analysis of variance (ANOVA) to test whether the perceived numerical estimates in the different verbal quantifier conditions (e.g., "rare," "very common") actually do line up with expectations. A Tukey post-hoc test was used to determine

the significance of specific intergroup differences. Hypothesis 3 was analyzed using several analyses of covariance (ANCOVA). The covariates of these ANCOVAs included the two numeracy measures, participants' perceptions of painkiller danger and addictiveness, and reported frequency of their seeing news coverage about and talking with friends and family about painkiller abuse. Pearson correlations were used to establish relationships and directionality. Hypothesis 4 was analyzed with a chi-squared test to determine whether participants in the high-severity conditions would have better recall of the verbal quantifier in the description. Finally, the study's research question was analyzed with an ANOVA to determine whether participants reported higher numerical estimates for either the *high severity* or *low severity* conditions, and the *rare* and *common* conditions.

RESULTS

Hypothesis 1

Hypothesis 1 was split into three parts regarding participants' numeracy. Hypothesis 1a posited that participants with higher numeracy scores would report higher confidence in their numerical interpretations of side effect likelihood, based on the medication descriptions. The first step in this analysis was to create two different variables, one by taking the mean scores for the objective numeracy scale and another by taking the mean scores for the subjective numeracy scale. Since the objective numeracy scale included eight open-ended responses and three multiple-choice questions, each of these questions was recoded into a binary correct-incorrect variable. This was done by looking through the open-ended responses and deciding which answers would be considered acceptable (e.g., "2%," "2 percent," "two percent") and recoding them as correct (i.e., 1) and recoding all other answers as incorrect (i.e, 0). The mean of these answers was then converted into a single variable to assess overall objective numeracy. Similarly, a single variable was created by taking the mean of the 8 subjective numeracy questions, which were each measured using a six-point scale. One of these questions was recoded, so that for each of the measures, a higher score indicated higher numeracy. The Chronbach's alpha for the subjective numeracy scale was .855, while it was .817 for the objective numeracy scale. This indicates that each of these scales has a high internal reliability, consistent with previous studies. The correlation between the two scales was 0.212, which is less than the 0.53 which has been reported previously (Fagerlin et al., 2007).

		Confidence	Estimate	Objective	Subjective
		in estimate	range	numeracy	numeracy
Confidence in estimate	Pearson Correlation	1	141**	251**	.140**
	Sig. (2-tailed)		.000	.000	.000
	Ν	724	702	712	701
Estimate range (high	Pearson Correlation	- .141 ^{**}	1	$.080^{*}$	098*
and low difference)	Sig. (2-tailed)	.000		.036	.011
	Ν	702	705	693	682
Objective numeracy	Pearson Correlation	251**	.080*	1	.212**
	Sig. (2-tailed)	.000	.036		.000
	Ν	712	693	716	699
Subjective numeracy	Pearson Correlation	.140**	098*	.212**	1
	Sig. (2-tailed)	.000	.011	.000	
	Ν	701	682	699	704

Table 4: Correlations between numeracy, confidence, and range

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Each of these numeracy measures was then correlated with participants' responses to a self-report question about the level of confidence in their quantitative estimation, as well as the range between the high and low estimates they gave (as a proxy for estimate certainty). The correlation tables for this analysis are shown in Table 4. The results showed that higher subjective numeracy scores were correlated with higher confidence in participants' estimates for both quantitative measures. For the numerical risk estimate, subjective numeracy had a Pearson correlation of 0.14 (p < .001). Objective numeracy had a Pearson correlation of -0.27 (p < .001) for the numerical risk estimate. This would suggest that individuals who scored highly on the subjective numeracy measure tended to have more confidence in their estimates, while the objective numeracy measure actually tended to predict a lack of confidence (expressed in the question as feeling "unconfident" about their estimate). Therefore, this hypothesis was supported by only the subjective measure of numeracy, whereas the objective measure did not support this hypothesis.

Hypothesis 1b predicted that higher numeracy scores would be correlated with a narrower range between participants' high and estimates for the second quantitative measure, as a proxy measure for certainty in their estimation. For this analysis, an Estimate Range variable was created by subtracting participants' low estimates from their high estimates. The Pearson correlation between subjective numeracy and range was -0.098 (p = 0.01), indicating that higher scores on subjective numeracy were related to a narrower range between participants' interpretations of the high and low estimations for the quantitative measure. The Pearson correlation between objective numeracy and range was 0.084 (p = 0.026), indicating that higher scores on the objective measure were related to wider ranges between the high and low quantitative estimates. Therefore, this hypothesis was supported by only the subjective measure of numeracy, whereas the objective measure did not support this hypothesis.

Hypothesis 1c predicted that subjective numeracy would be a better predictor of quantitative estimate confidence than objective numeracy. This was analyzed by comparing the relationships between the three variables. As the above findings suggest, each of the measures was significantly predictive of the confidence estimates, but in different directions. Since the subjective numeracy score was predictive of greater confidence than the objective measure, Hypothesis 1c was supported by the data.

Hypothesis 2

Hypothesis 2 stated that participants' estimations of the side effects' likelihood would increase in the order of the verbal quantifiers, such that *very rare* would have the lowest mean estimation and *very common* would have the highest. As a control, one condition for each of the

high and low severity messages did not contain a verbal quantifier at all, but instead said simply that "side effects include _____".

This hypothesis was tested by using an ANOVA to compare the mean estimates across the different conditions. The results supported the hypothesis, with the estimations increasing as the verbal quantifiers increased. Participants interpreted a very rare side effect as affecting an average of 208.97 (SD=256.89) out of 1,000 users of the medication, with rare side effects affecting an average of 234.54 (SD=263.10) users. *Common* was quantified as affecting an average of 334.83 (SD=286.37) out of 1,000 users, while very common side effects were believed to afflict 426.10 (SD=285.54) medication-takers. The control ("include") condition, which did not have a verbal quantifier at all, was seen to be a midpoint, with participants reporting this likelihood as 305.47 (SD=257.13). The F score for the differences between the conditions was $F_{4,707}= 14.44$, p < .001, eta²= 0.08. A Tukey post-hoc test showed that the majority of the conditions were statistically significantly different from each other, with several exceptions. Very common was significantly different from common (p = .037), rare (p < .001), *very rare* (p < .001), and *include* (p = .002). *Common* was significantly different from *rare* (p = .002). .016) and very rare (p = .001). Very rare was significantly different from *include* (p = .021). *Include* was not significantly different from either *common* (p = .890) or *rare* (p = .179), while *rare* was also not significantly different from *very rare* (p = .930).

Hypothesis 3

Three ANCOVAs were conducted to analyze the effects of several covariates on the three risk perception measures. The ANCOVA tables for the numerical estimate, side effect likelihood, and side effect worry DVs are shown in Table 5, Table 6, and Table 7, respectively.

Hypothesis 3a predicted that participants who felt that prescription painkillers were dangerous would perceive greater risk from the listed side effects in the message they read. This was done by reverse coding responses to the question statement "prescription pain medications are dangerous to those who take them," so that higher scores indicated more agreement, and looking at both the correlations between this variable and the three (*gist* and *verbatim*) risk perception measures, as well as the results of an ANCOVA for each of the outcome variables. The Pearson correlation between perceptions of painkiller danger and the numerical risk estimate was .142 (p < .001). For the *gist* measures, the correlation between perceptions of painkiller danger and participants' belief that they would be likely to experience side effects was .417 (p < .001), while the correlation with participants worry about the side effect was .345 (p < .001). An ANCOVA revealed that the relationship between perceptions of painkiller danger and the numerical estimate was $F_{1.678}$ =.458, p=.499, eta²=.001.

The relationship between perceptions of pain medication danger and participants' *gist* perceptions that they would be "likely to experience side effects" was $F_{1,687}$ =49.882, *p*<.001, eta²=.069. The relationship between their perceptions of pain medication danger and whether they would be worried about side effects was $F_{1,689}$ =30.073, *p*<.001, eta²=.043. This suggests that there is a fairly strong and highly statistically significant relationship between the perception of painkiller danger and the *gist* measures of risk. This signals that participants' baseline perceptions of prescription pain medication danger had a larger influence on their *gist*

interpretations of risk than their verbatim interpretations. Therefore, the hypothesis was

supported for two of the three risk measures.

Table 5: Analysis of Covariance for Numerical Estimate

-	Type III Sum					Partial Eta
Source	of Squares	df	Mean Square	F	Sig.	Squared
Corrected Model	8636109.921 ^a	15	575740.661	8.436	.000	.160
Intercept	1114586.588	1	1114586.588	16.331	.000	.024
PPMs are dangerous ¹	31250.473	1	31250.473	.458	.499	.001
PPMs are addictive ¹	277253.849	1	277253.849	4.062	.044	.006
Frequency of news coverage about PPMs ¹	56646.723	1	56646.723	.830	.363	.001
Frequency of speaking with friends/family about PPMs ¹	370226.128	1	370226.128	5.425	.020	.008
Objective numeracy ¹	1872135.924	1	1872135.924	27.431	.000	.040
Subjective numeracy ¹	25995.493	1	25995.493	.381	.537	.001
Severity ²	21321.569	1	21321.569	.312	.576	.000
Verbal quantifier ²	3939141.938	4	984785.484	14.429	.000	.080
Severity * Verbal quantifier	213597.260	4	53399.315	.782	.537	.005
Error	45180519.330	662	68248.519			
Total	115592556.00	678				
	0					
Corrected Total	53816629.250	677				

Dependent Variable: Numerical estimate*

a. R Squared = .160 (Adjusted R Squared = .141)

* "If 1,000 patients took Soulagis (the drug you just read about), around **how many** of these patients would you expect to experience the listed side effect?"

¹ Covariate

² Factor

Hypothesis 3b predicted that participants who believed that prescription painkillers had a high potential for addiction and abuse would have higher perceptions of risk from side effects. The Pearson correlation between belief in painkillers' potential for addiction and the numerical risk estimate was $0.158 \ (p < .001)$. For the *gist* measures, the Pearson correlation between belief about addiction and participants' belief that they would be likely to experience the listed side effect was $0.34 \ (p < .001)$, while the correlation with worry about side effects was $.301 \ (p < .001)$. This indicates that the more strongly a participant believed that prescription pain medications had a high potential for addiction, the higher they estimated the likelihood of the listed side effect.

Relationships were also analyzed with the three ANCOVA results for the different risk perception measures. The relationship between participants' belief that prescription pain medication users were likely to become addicted and the numerical estimate was $F_{1,678}$ =4.062, p=.044, eta²=.006. While the latter relationship was significant, these results suggest that baseline beliefs about the addiction potential of prescription pain medications does not play a large role in their *verbatim* risk interpretations.

The relationship between the perceptions of prescription pain medication users' potential for addiction and participants' beliefs that they would be likely to experience the listed side effects was $F_{1,687}=10.810$, p=.001, $eta^2=.016$. The relationship between perceptions of pain medications' addiction potential and participants beliefs that they would be worried about the listed side effect was $F_{1,688}=13.165$, p<.001, $eta^2=.019$. This suggests that not only do baseline beliefs about prescription pain medications' potential for addiction relate highly with *gist* measures of side effect risk perceptions, but that they are much less related to the *verbatim* measures.

Table 6: Analysis of Covariance for Perceptions of Side Effect Likelihood

	Type III					
	Sum of		Mean			Partial Eta
Source	Squares	df	Square	F	Sig.	Squared
Corrected Model	191.010 ^a	15	12.734	15.085	.000	.252
Intercept	60.555	1	60.555	71.735	.000	.097
PPMs are dangerous ¹	42.108	1	42.108	49.882	.000	.069
PPMs are addictive ¹	9.125	1	9.125	10.810	.001	.016
Frequency of news	.906	1	.906	1.074	.301	.002
coverage about PPMs ¹						
Frequency of speaking	1.084	1	1.084	1.284	.258	.002
with friends/family about						
PPMs ¹						
Objective numeracy scale ¹	13.663	1	13.663	16.186	.000	.024
Subjective numeracy scale ¹	.035	1	.035	.041	.840	.000
Severity ²	2.403	1	2.403	2.847	.092	.004
Verbal quantifier ²	24.967	4	6.242	7.394	.000	.042
Severity * Verbal	4.794	4	1.199	1.420	.226	.008
quantifier						
Error	566.426	671	.844			
Total	7281.000	687				
Corrected Total	757.435	686				

Dependent Variable: Side effect likelihood*

b. R Squared = .252 (Adjusted R Squared = .235)

* "If I took this drug, I would be likely to experience side effects"

¹ Covariate

² Factor

Hypothesis 3c predicted that participants who had seen more news coverage (the type of medium was not specified) about painkiller abuse and addiction would be more likely to have greater risk perceptions. This was analyzed by comparing the correlations of viewership of news coverage with risk perception measures. The Pearson correlation between news coverage and the numerical risk estimate was 0.039 (p = .296). For the *gist* measures, the Pearson correlation between news coverage and participants' belief in the likelihood of their experiencing side effects was .062 (p = .095), while it was .074 (p = .049) for their worry about side effects. These results were also analyzed with an ANCOVA. The relationship between participants' viewing news coverage and the numerical risk estimate was $F_{1.678}$ =.830, p=.363, eta²=.001. This indicates a weak and statistically insignificant relationship between participants' viewership of news coverage about prescription pain medications and *verbatim* measures of their risk perceptions from the listed side effects.

The relationship between news coverage of prescription pain medication abuse and participants' perception that they would be likely to experience the listed side effects was $F_{1,687}=1.074$, p=.301, eta²=.002. The relationship between watching news coverage and participants' reporting that they would be worried about the listed side effect was $F_{1,689}=.448$, p=.504, eta²=.001. This indicates quite a weak and statistically insignificant relationship between participants' reported viewing of news coverage about prescription painkiller abuse and how worried they were about the listed side effects in the message.

Table 7: Analysis of Covariance for Perceptions of Side Effect Worry

Dependent Variable: Side effect worry*							
	Type III Sum					Partial Eta	
Source	of Squares	df	Mean Square	F	Sig.	Squared	
Corrected Model	304.796 ^a	15	20.320	16.823	.000	.273	
Intercept	60.256	1	60.256	49.886	.000	.069	
PPMs dangerous ¹	36.324	1	36.324	30.073	.000	.043	
PPMs addictive ¹	15.902	1	15.902	13.165	.000	.019	
Frequency of news coverage about PPMs ¹	.541	1	.541	.448	.504	.001	
Frequency of speaking with friends/family about PPMs ¹	4.020	1	4.020	3.328	.069	.005	
Objective numeracy ¹	.507	1	.507	.420	.517	.001	
Subjective numeracy ¹	4.988	1	4.988	4.130	.043	.006	
Severity ²	129.933	1	129.933	107.573	.000	.138	
Verbal quantifier ²	7.539	4	1.885	1.560	.183	.009	
Severity * Verbal	7.399	4	1.850	1.531	.191	.009	
quantifier							
Error	812.887	673	1.208				
Total	8085.000	689					
Corrected Total	1117.684	688					

Dependent Variable: Side effect worry*

c. R Squared = .273 (Adjusted R Squared = .256)

* "If I took this drug, I would be worried about experiencing side effects"

¹ Covariate

² Factor

Hypothesis 3d predicted that participants who had spoken more with friends and family about prescription painkiller abuse and addiction would be more likely to have greater perceptions of risk from the drug's side effects. This was first analyzed by looking at the correlations between frequency of speaking with friends and family about painkiller abuse and the three risk perception measures. The Pearson correlation for speaking with friends and family and the numerical risk estimate was .172 (p < .001). For the *gist* measures, the Pearson correlation between talking about prescription painkiller abuse and participants belief in the likelihood of experiencing side effects was .187 (p < .000), while it was .153 (p = .001) for participants' worry about the listed side effects. ANCOVAs revealed that the relationship between participants' speaking with friends and family about prescription painkiller abuse and their response to the single-estimate quantitative measure was $F_{1,678}=5.425$, p=.020, eta²=.008. This indicates that the relationship between participants' amount of speaking with family and friends about painkiller abuse was significantly related to the size of their numerical estimate.

The relationship between participants' conversations with friends and family about prescription pain medication abuse and their beliefs that the medication they would be likely to experience the side effects from the listed drug was $F_{1,687}=1.284$, p=.258, eta²=.002. The relationship between participants conversations and their level of worry about side effects from the described drug was $F_{1,689}=3.328$, p=.069, eta²=.005. This indicates that the amount that participants spoke with friends and family about prescription pain medication abuse had a relatively weak relationship with the *gist* measures of their risk perceptions.

Hypothesis 4

Hypothesis 4a predicted that participants in the high-severity conditions would have a better recall of the exact verbal quantifier used in the side effect description than those in the low-severity conditions. The findings suggest that although this was the case, it was not statistically significant. In the low-severity condition, 25.8% of participants were able to correctly recall the exact verbal quantifier they were given (i.e., the adverb-adjective word pair). For the high-severity condition, 31.8% of participants correctly recalled the description. While these were in the predicted direction, these findings were not significant. The chi-squared

statistic for this was 3.21, with a statistical significance of p = 0.073. Therefore, this hypothesis was not supported at a statistically significant level.

Hypothesis 4b predicted that participants in the *rare* conditions would have a better recall of the exact verbal quantifier used in the description than those in the *common* conditions. The results revealed that participants in the *rare* conditions had much greater recall (42.8% correct) than those in the *common* (16.5%) conditions, or the control (25.2%), which did not have a verbal quantifier. The chi-squared statistic for this difference was 50.36, with a statistical significance of p < .001. Therefore, hypothesis 4b was supported.

Research question 1

Research Question 1 asked if participants would report higher numerical estimates for the *high severity* or the *low severity* side effects, since the literature has been split on this. This was tested by conducting an ANOVA on the numerical estimate, compared by the side effect severity. The results showed that although participants gave somewhat lower estimates for the *high severity* side effect than for the *low severity* side effect, the results were not statistically significant. In their numerical risk estimates, participants gave a mean estimate of 307.03 for the *low severity* conditions and 295.85 for the *high severity* conditions. This difference did not approach statistical significance, however, with a p = .595. Therefore, while the estimates for low severity were slightly higher, there was not a meaningful difference between the two conditions.

DISCUSSION

This experimental study assessed how participants interpreted and quantified the risks from side effects of a prescription pain medication based on a risk message, as well as what sorts of underlying factors influenced these interpretations. Grounded in fuzzy trace theory, this study explored the effects that verbal quantifiers, side effect severity, numeracy, and background perceptions of pain medications had on participants' risk perceptions of a new medication's side effects. Results yielded a number of interesting findings, such as differences in the predictive power of subjective and objective numeracy, the effects of baseline beliefs about pain medications on participants' interpretation of new information about side effect risk, and the differences between *gist* and *verbatim* measures of risk.

The following sections of this chapter discuss the importance and potential significance of these findings, thematically. First, the startling differences between numerical estimates and experts' recommendations for how these terms should be interpreted will be explored, before an investigation of the effects of participants' preconceptions of pain medications on their *gist* and *verbatim* interpretations. Then, it will discuss the intriguing findings surrounding the objective and subjective measures of numeracy, as well as considering the effects of type of message on participants' precise recall of the verbal quantifier used. The dissertation will conclude with the contributions that this study makes the theory and future research, as well as practice and policy, before discussing the potential limitations to this study's generalizability to other contexts.

Verbal quantifiers and participants' numerical estimates

One of the most immediately surprising findings of this study was how high participants' numerical estimates were—for all of the conditions. The second hypothesis stated that

participants' quantitative estimates would be in the order of the verbal quantifier used, such that they would give the smallest estimates for the "very rare" message and the largest for the "very common" message. While the results did support this order of interpretations, this study bolsters the findings of other studies (Cox, 2016; Berry, Knapp, & Raynor, 2002; Berry, Raynor, & Knapp, 2003) in showing that participants' estimates were dramatically higher than the European Union recommendations for how these terms should be interpreted. As a reminder, Table 8 shows the EU recommendations for how consumers should interpret these terms. Note that, out of 1,000 users of a medication, a "rare" side effect is posited to affect no more than one out of those 1,000, while a "common" side effect should affect between 10 and 100 users. The comparison with the findings from this study's quantitative estimate is quite striking.

While a "very rare" side effect in the EU recommendations is supposed to affect between .001% and .1% of users, participants in this study gave a mean score of 208.97, or nearly 21% of users being affected. This is clearly a much larger interpretation of these terms than would be expected, given experts' recommendations. Of course, the low end for the EU recommendations was impossible for participants to truly report in this study because the sliding scale only represented 1,000 individuals—and a side effect can't affect .1 people. However, the values of the sliding scale were limited to 1,000 because 10,000 would likely be too large of a number for what amounted to several inches of a scale on participants' computer screen. In other words, 1,000 was a compromise between having enough possible values that it could be meaningfully analyzed (e.g., a possible estimate range of 1-100 would take out a lot of the nuance in individuals' responses) but also so that participants could have some precision in their estimates and not have a millimeter represent hundreds of points.

Verbal Quantifier	Recommended range of interpretations				
"Very common"	> 10% of users affected				
"Common"	1% to 10% of users affected				
"Uncommon"	.1% to 1% of users affected				
"Rare"	.01% to .1% of users affected				
"Very rare"	.001% to .01% of users affected				
"Exceedingly rare"	< .001% of users affected				

 Table 8: EU Recommendations for Verbal Quantifiers of Side Effect Risk

These findings beg the question of why there is a such a wide gap between the EU experts' recommendations of how these kinds of verbal quantifiers are *intended* to be interpreted, and how members of the public *actually* interpret them. Some of the factors in this study (and those like it) that may cause participants to inflate their estimates (e.g., an advertising source vs. a physician source, emotionally charged reactions to pain medications in particular) are discussed in the limitations section at the end of this dissertation. However, it could also be (as it often is) that the simplest explanation is the best one. That is, the individuals who are creating and endorsing these interpretive recommendations are physicians and pharmacologists—people with much more intimate and specialized knowledge of medications and side effects than members of the general public. As such, they have a much more informed and relevant frame of reference for the relative likelihood of side effects occurring. For example, participants in this study estimated that a *common* should represent an incidence of between 1% and 10%. Since those with specialized and intensive knowledge of medication and pharmacology are sure to

understand that the vast majority of drug side effects are unlikely to affect huge numbers of users, their interpretation of a *common* side effect is sure to be lower estimate. Members of the public, who will not have such a background, are likely to use more general applications of what the words *common* and *rare* mean to them, from everyday parlance (e.g., "it's rarely this cold in May").

Preexisting drug perceptions, and gist and verbatim processing

One of the major topics that this study sought to explore was the differential impact of preconceptions about prescription pain medications on participants' gist and verbatim traces of risk information from the study's side effect message. The study's third set of hypotheses focused on the effects that existing perceptions about prescription pain medications' danger and potential for addictiveness, how often participants saw news coverage of painkiller abuse, and how often they spoke with friends and family about abuse, had on their perceptions of risk from the new drug's side effects. Analyses of these results reveal interesting findings, in particular when it comes to the propositions of fuzzy trace theory. Importantly, both of the general, background risk perceptions—about prescription drugs' danger and addiction potential appeared to have significant influences on participants' risk perceptions. This is supportive of the idea that individuals do not just interpret risk information in isolation, but instead actively draw upon their more general, underlying beliefs when interpreting new information. Thus, individuals' reactions to risk communication messages are not simply a stimulus-response, but are highly influenced by existing beliefs. In some cases, they may be more influential than the message itself. Note from ANCOVA 2 that participants' interpretation of the likelihood of being affected by side effects from the drug described in the experiment had more of its partial

variance explained (eta²) by the underlying belief in pain medications' danger than either of the two manipulations—side effect severity or verbal quantifier.

There is also important evidence in these results for one of fuzzy trace theory's main tenets—that *gist* processing is inherently more influenced by individuals' background, existing perceptions than *verbatim* processing is. This was found to be the case here, as the baseline beliefs about pain medications' danger and addiction potential were highly significantly related to the *gist* risk perception measures. On the other hand, addiction concerns had only a very modest effect on participants' numerical risk estimates, while overall perceptions of pain medications' danger did not have a significant relationship at all with the numerical measure. The discrepancy between these covariates' effects on the outcome measures would seem to indicate that these types of measures are inherently different from one another in terms of participants' processing of risk. Further research should be conducted in this area to better understand the underlying influences of different types of risk interpretations and the ways in which they can be measured. The initial results presented here appear to support fuzzy trace theory's assertions that different types of processing of risk information can occur concurrently, with preexisting beliefs having different effects on these interpretations.

Recall that fuzzy trace theory proposes a difference between verbatim processing (or a direct and literal interpretation of new information) and gist processing (or a bottom-line interpretation of new data, integrated with existing beliefs, experiences, and attitudes). In order to better reflect these differences, three primary measures were utilized to assess participants' perceptions of risk from the side effects of the drug described in the experimental message. One of these asked respondents to give a direct quantitative estimate, which was used as a verbatim measure of risk. This was because it required participants to translate a verbal description of the

pain medication's side effects to a direct, numerical estimation. By asking for a specific, singular numerical estimate, participants were prompted to consider potential risk in strictly numerical terms. In other words, the aim of the question was to prompt them to consider the quantitative risk based more singularly on the message itself, and not to include other background factors in their estimate.

The two other measures were also employed to assess risk perceptions, but this time from a more overall, bottom-line (gist) interpretative perspective. The first of these measures employed Likert-style responses to a question that asked about the "likelihood" of being affected by side effects from the drug. The more general nature of this question aimed at capturing participants' overall beliefs of their likelihood of being affected by this drug's side effects. It was designed to be more general in scope, in order to capture existing, baseline beliefs in their interpretations. The belief was that a more general measure of risk would be more likely to be affected by beliefs about issues such as prescription pain medications' overall danger to users, and the likelihood that individuals who take them would be highly susceptible to addiction.

The third measure asked for a Likert-style response to a question about participants' overall level of *worry* about the side effects of the medication. This was considered to be the most general risk measure because it separated itself from just the quantification and also invited the most integration with participants' baseline beliefs. In other words, participants' pre-existing beliefs about prescription pain medications would be more likely to influence responses to this measure, because it asks about the level of *worry*, which is a more general concept. For example, many people may believe that a medication that has the potential side effect of intense nausea and vomiting would be very worrisome, regardless of how likely they believe it is to occur. Some studies have shown just such an effect, where perceptions of severity play a huge

role in medical worry, whereas perceptions of the likelihood of being affected play a relatively small one (Loewenstein et al., 2001). The difference between these types of measures should allow for a parsing out of appropriate ways to assess individuals processing of risk information in either verbatim or gist ways.

Numeracy—objective and subjective measures

One of the more surprising findings from this study related to the differences between the subjective and objective numeracy scales. The first set of hypotheses predicted that higher scores on the subjective and objective numeracy scales would be related to both a higher confidence in participants' numerical estimates, and a narrower range (higher certainty) in their high and low estimates. It also predicted that subjective numeracy would be a greater predictor of participants' confidence and certainty (based on an earlier version of the present study). The results of these hypothesis tests revealed differences between the relationships of subjective and objective numeracy with participants' confidence and certainty (or the narrowness of the highlow range) in their numerical estimates. While a higher score on the subjective numeracy scale was found to be related to higher confidence in participants' numerical estimates, a higher score on the objective numeracy scale predicted a lack of confidence in a participant's estimate. Since the subjective numeracy scale asks for self-reports, it was expected that this would be related to higher confidence (and a narrower range of estimates), just as it was expected that a scale that more objectively measures mathematical skill would also be related to higher scores in both. Therefore, it was rather unexpected to find that subjective and objective numeracy had not only differing effects, but completely opposite effects on both the self-report measure of confidence and the estimate-range measure, which was used as a proxy here for estimate certainty.

In an earlier study (Cox, 2016), only the measure of the range of low and high estimates was analyzed in relation to numeracy. Similar findings on the opposite effects between the numeracy scales reflect what was found in the present study. Upon further reflection of the findings of Cox's earlier study, some doubt was cast upon the idea that a wider range between low and high estimates could actually be considered a true measure of one's certainty in their estimate. It may very well be that it is more accurately a measure of something else—such as the understanding of their limitations in estimating risk. Because of this possibility, a more direct question was used here that asked participants to directly report how confident they were in the single numerical estimate they had given. The finding that subjective and objective numeracy were found to be, respectively, positively and negative predictive for both the confidence and certainty measures lends some credibility to the idea that these latter two measures are actually assessing similar constructs. In other words, while the confidence measure asks for a direct selfreport, and the certainty measure is a more direct analysis of the range of their estimates, the results indicate that these constructs are highly similar. This is supported by the fact that confidence and certainty were correlated at -.141 (p < .001), meaning that higher confidence was related to a narrower range between high and low estimates.

This leads to the question of why two different ways of measuring what is ostensibly the same construct—an individual's ability to understand and utilize mathematical concepts—would not only reveal different relationships with individuals' confidence in making numerical estimates, but completely opposite (and statistically significant) relationships. This new result requires a greater investigation into the nature of these two types of measurement.

While the objective measures are akin to an actual math test, directly asking participants to make mathematical calculations on the spot, the subjective numeracy scale asks respondents to

self-report their own ability with and preferences for using numbers. On the face of it, these measures might appear to be of different constructs entirely. The objective scale clearly appears to assess whether participants' responses were correct or incorrect (insofar as they actually respond to the question and don't leave it blank), it could be that the subjective numeracy scale is more accurately a measure of participants' numerical self efficacy, as opposed to an assessment of their actual skill. After all, an individual's strong feelings about their ability to do a particular thing well do not necessarily translate into better performance in that area.

In some ways, this is similar to the Dunning-Kruger effect, where individuals who have low or marginal ability in certain areas often tend to vastly overstate perceptions of their own personal abilities. Across several studies assessing college students' self-reports of confidence in several areas and then direct assessments of their competence, researchers found that low-ability students tended to greatly inflate how well they thought that had scored in several different areas. The psychologists found that "participants scoring in the bottom quartile on tests of humor, grammar, and logic grossly overestimated their test performance and ability. Although test scores put them in the 12th percentile, they estimated themselves to be in the 62nd" (Kruger & Dunning, 1999, pp. 1130)

In fact, some studies have shown that there is an inverse relationship between belief in one's math ability and their actual math ability (IEA, 2009). For example, studies that have rated and compared academic ability in middle school students across the world have asked students to report how good they are at math before administering a direct quantitative ability assessment. In some cases, American students (and more generally those from Western countries) have tended to rate their math skills very highly and then earned mediocre scores on the actual assessments (IEA, 2009). American students' poor showing in international comparisons has

been shown time and again in studies. In 2017, a Pew research report found that American 15year-olds placed 38th among those from 71 countries studied (Pew Research Center, 2017). Therefore, these findings may be more widely indicative of poor numeracy among Americans, more generally.

These findings may suggest that the subjective numeracy scale would be better described as a measure of mathematical self-efficacy, than as a measure of individuals' actual mathematical ability. It is also worth noting that both the subjective and objective numeracy measures are highly statistically significant predictors of confidence and certainty. Therefore, it is not as if either of these measures is an irrelevant factor when considering participants' confidence or certainty in their estimates.

So why then do these measures appear to be very significant predictors of starkly opposite effects? It could have something to do with this hypothesis' specific focus on the construct of confidence/certainty. In other words, these relationships between numeracy and confidence may be different than for other outcome measures. In order to test this idea, Pearson correlations were also analyzed between subjective and objective numeracy and individuals' ability to correctly recall the exact verbal quantifier they had been presented in the medication description. This analysis revealed that the correlation between recall and subjective numeracy was 0.135 (p < .001), while recall and objective numeracy were correlated at 0.177 (p < .001). While objective numeracy has a slightly higher correlation than subjective numeracy, higher scores on both of these measures were related to better recall of the exact verbal quantifier used.

This possibly indicates that participants' confidence or certainty in their estimates may be a unique situation in which these two measures would give opposite predictions. In the future, studies should look at the relationships of both types of numeracy with other variables, to see if

there are other situations in which the scales present highly statistically significant, but completely opposite relationships.

Severity, common/rare, and recall of verbal quantifiers

One set of hypotheses also focused on participants' ability to recall the exact verbal quantifier (or the control condition, which did not feature one) that had been used in the description they had been presented in the description of the medication. The two parts of this analysis focused respectively on the effects of severity and whether the verbal quantifier had been in the rare ("rare" or "very rare") or common ("common" or "very common") message conditions. Results showed that although participants who were in the high-severity condition (31.8%) tended to have better recall of the term used than those in the low-severity condition (25.8%), this difference was not a statistically significant one. This was a somewhat surprising finding, since it was believed that individuals would find the high-severity side effect to be more striking (and therefore more memorable), and would lead participants to remember the likelihood of the effects better. After all, if there's a side effect that's particularly worrying, it's reasonable to expect that participants would pay more attention to the likelihood that it would occur. An induction check in both the pre-test and the main study confirmed that participants did, in fact, pay attention to the severity, so this is not simply individuals ignoring this aspect of the message. In light of this, the lack of effect is curious and should be explored with further research on this topic.

The second part of this hypothesis stated that individuals who were presented with a message description that included the word "rare" would have better recall than individuals in the "common" conditions. This was largely due to the results of Cox' (2016) preliminary study,

which indicated that people tended to pay more attention to the messages in the "rare" conditions. The findings suggested that this was the case in the present study, since individuals in the *rare* conditions tended to have much higher recall (42.8%) of the exact verbal quantifier than those in the *common* (16.5%), or control (25.2%) conditions. The large difference between these main conditions was both striking and statistically significant (p < .001). Participants in the *common* conditions had even lower recall than those in the control conditions. Conclusions about the control condition should likely be taken with a grain of salt, however, since it is very likely easier to remember that there was no verbal quantifier at all, than to remember the exact adverb/adjective pair that was included in the message.

One possible explanation for individuals' better recall of the rare conditions is that "rare" likely has a narrower range of possible interpretations than "common." In other words, participants may perceive a more significant difference between two events described as "rare" and "very rare" than they do those described as "common" and "very common." This is also reflected in the EU's recommendations for these term's interpretations, since the range of recommended interpretations for each of the "common" conditions is much higher than for the "rare" ones. For example, according to these recommendations, a "rare" side effect is to be interpreted as affecting between .01% to .1% percent of users, while a "common" side effect should occur in between 1% to 10% of users of the medication. This means that the range of "acceptable" interpretations for a "common" side effect is much larger than those for a "rare" side effect.

Out of a group of 10,000 people taking the medication, the CIOMS recommendations indicate that a "rare" side effect would occur in between 1 (i.e., .01% of 10,000) and 10 (i.e., .1% of 10,000) of individuals. On the other hand, a "common" side effect would occur in between

100 (i.e., 1% of 10,000) and 1,000 (i.e., 10% of 10,000) users. This represents a possible range of interpretations of 9 for the term "rare" and 900 for the term "common." In other words, "common" is being used to describe an acceptable range of interpretations that's nine times as big as "rare." This is a startlingly big difference that may not be immediately apparent when glancing at the list of ranges.

The study's sole research question asked if participants would report higher numerical estimates for the *high-severity* or *low-severity* side effects. Analyses showed that participants reported higher estimates for the low severity side effects, when compared with the high severity ones, but there was no significant difference between them. Given the fact that participants appeared to perceive high- and low-severity side effects very differently, with the induction checks in the pre-test and main study, this is not simply a result of there being no difference between the two. Rather, it would suggest that perceptions of a side effect's severity did not appear to have a meaningful impact on how likely the side effect was believed to occur. This is an intriguing finding, given the conflicting literature on the issue. It could be that individuals simply separate the two issues, such that severity has a large impact on their worry, but relatively little impact on their perceptions of likelihood.

Implications for theory and research

Part of the purpose of this study was to explore some of the concepts of fuzzy trace theory. In particular, the three different measures of risk perceptions were designed to assess individuals' gist and verbatim processing by comparing the differences between their direct interpretations of the risks presented in the medication's message and participants' more general, underlying evaluations of the medication's risks. By attempting to directly measure these

different types of processing of the same information, these findings contribute not only to the validity of the theory, but also to future studies that may try to further parse out and measure different types of risk information processing.

While many studies that assess individuals' reactions to a message look only at the message's direct effects, or include demographic covariates such as education and age, this study aimed to take a more holistic approach to studying the perceptual influences of individuals' risk interpretations of a message. Since fuzzy trace theory states that gist impressions will be influenced by previous beliefs and experiences, as well as other factors, such as numeracy, these kinds of measures seemed natural for inclusion in this study. The finding that preexisting beliefs about prescription pain medications' inherent danger and potential for addiction had a greater influence on individuals' more general risk perceptions than their verbatim ones was therefore an important aspect for the improved understanding of underlying beliefs of *gist* vs. *verbatim* processing of risk information.

There are a number of findings here that should suggest work for researchers to explore further in the future. In particular, researchers should continue to develop measurements for different types of risk information processing (i.e., *gist* and *verbatim* traces). Since different baseline beliefs clearly have different effects on individuals' gist and verbatim interpretations of risk, this is an important avenue for future research. Other kinds of potential baseline influences should also be included in studies to examine their impact on perceptions of risk (e.g., perceptions of prescription drugs more generally, worries about prescriptions being under-tested before being released, personal experiences taking the drugs, etc.)

Additionally, the two measures of numeracy are an interesting case, where some outcome measures showcase a clear difference in their predictive effects, while others demonstrate their

ability to similarly predict outcome variables. Therefore, these two different types of scales should be more carefully explored in terms of their relationships with different measures of risk perceptions. In particular, other kinds of outcome measures that may have more to do with the understanding of math and numbers could be useful for better parsing out the differences in subjective and objective scales.

Implications for practice and policy

Although verbal quantifiers are used extensively in direct-to-consumer advertisements, individual's interpretations of what these terms mean are inconsistent. For each of the verbal quantifier conditions in this study, the standard deviations of the numerical estimates were higher than 250. For a 1,000-point scale, this is quite high, suggesting a large amount of variation in individuals' responses. These findings indicate that these terms are rather blunt instruments for the conveyance of quantitative risks. What, then, does this mean for policymakers and pharmaceutical companies? At the very least, it would suggest that pharmaceutical companies rethink their choice to utilize these terms in advertisements. A seemingly reasonable solution may be to suggest that they abandon the use of these terms altogether, opting instead to report more precise statistics, such as "in clinical trials, 3% of participants experienced dry mouth."

While this may seem to be a natural implication of this study (and others that suggest these terms aren't effective), putting this into practice would very likely be difficult. First, if there are no FDA regulations for how these risks must be communicated, then pharmaceutical companies are unlikely to change anything about the way they communicate. More centrally, however, is the fact that specific numbers may not be practical information for consumers. Many side effects will be quite rare, occurring in a miniscule number of users. If ads are hyper-

precise, will this help users to really understand their risk likelihood? In other words, if an advertisement explains that a side is likely to occur in 0.001% of users, would consumers meaningfully differentiate between this and one that occurs in 0.01% of users (despite the latter being more likely, by a factor of ten?) It is even possible that consumers with low numeracy may downplay these risks, since they both may seem to them as incredibly small likelihoods.

Because the FDA does wield some significant power over companies' advertisements for prescriptions, it could be within their ability to mandate a standardization for terms such as verbal quantifiers. Evidence (from this study and others) suggests that consumers do not interpret these terms the same way as the EU guidelines suggest. This means that further studies that look into how individuals interpret these words are highly warranted. If agencies and companies can better understand how people interpret these terms, then they can more precisely communicate them.

A more general suggestion, which is likely to reap a number of benefits for all involved, would be for pharmaceutical companies and agencies such as the FDA to conduct further studies on how people interpret and respond to information regarding prescription drugs—and in particular prescription pain medications. One important contribution of the present study is that it demonstrates that it is not enough to simply present individuals with a message and then gauge their responses to it. Studies also need to incorporate participants' underlying beliefs and experiences that are relevant to prescription medications. A better understanding of the bigpicture mental approach (or schema) that individuals use when evaluating risk information about the drugs would no doubt help risk communicators and physicians to communicate risk more effectively.

Additionally, taking baseline beliefs as important factors in risk interpretations may help physicians to better talk with their patients about the risks and benefits of prescription medications. Baseline beliefs, experiences, and behaviors may be important risk factors that could affect the way they speak with patients about the potential for addiction and abuse. An understanding of these beliefs, and potentially patients' numeracy, could potentially help physicians to target communications to certain populations. The better communication of risks from prescription pain medications is something that benefits all stakeholders. Consumers of these drugs could benefit greatly from a better understanding of the potential risks and tradeoffs associated with taking these drugs. Pharmaceutical companies and federal agencies would also have much to gain, because less abuse of these drugs would make their jobs easier and lessen the blame they have taken for a failure to stop addiction problems nationwide.

All of these factors lead to a three-step set of recommendations that the author of the present study has for policy-makers, regulatory agencies, and pharmaceutical companies. The first is to work together to conduct studies on how members of the public actually interpret these terms. The use of these terms is of little benefit if there is such a wide gap between the interpretations of experts who came up with them and the individuals who will base their impressions on these terms. Before anything meaningful can be done to better communicate the meaning of these terms, interpretations of what these terms mean and how they are used in advertisements need to be much more closely aligned. Since this goal of better communication is something that would benefit all stakeholders, it should ideally be a collaborative series of projects between agencies such as the FDA, pharmaceutical companies, and academic researchers. Once there is a better understanding of how members of the public actual quantify and interpret these terms, actual recommendations can be more meaningful.

The next step is that the FDA should actually mandate the use of these terms in direct-toconsumer ads, so that there is a standardized set of guidelines for how the descriptions should be used and interpreted. This is unlike in the EU, where these terms' interpretations are recommendations and not actually mandated by any regulatory body. If the FDA does not require that these terms be standardized, pharmaceutical companies are unlikely to follow mere suggestions. At the time of this writing, the FDA has a fair amount of control over passing regulatory laws aimed at protecting members of the public. Therefore, they would have the ability to wield some power over the pharmaceutical industry in how these terms should be communicated and interpreted.

The final step is for these research-supported, mandated terms to be actually communicated to members of the public, so that they understand that that these terms have actually been standardized. There are several ways that this could be done in such a way that people can be exposed to it and better understand it. One potential method is to require that a list of terms and their standardized interpretations be included in print and television advertisements. In a print ad this could take the form of a small table on the "second page" of the advertisement (where more specific information is usually placed) that has the list of terms and their intended interpretations (e.g., *rare* = XX% and XX% of users.) This may help consumers to better navigate these advertisements.

Additionally, it may be useful to have an independent ad campaign from the FDA and/or pharmaceutical companies that is specifically about this change in these terms' standardization. Ads that focus on the fact that these terms now have standardized interpretations could go a long way in familiarizing members of the public in how commonly used words should be perceived in the context of a pharmaceutical advertisement. Since the better communication of side effect
risk information should be a common goal of regulatory agencies like the FDA (since this is a significant part of their purpose) and pharmaceutical companies (who should want individuals to use their products safely and appropriately) alike, this could be an important opportunity for collaboration towards a common, and important, public good.

Limitations

There are several issues that may limit this study's generalizability to wider populations. One limiting factor is the fact that this study only looked at one particular type of prescription medication—a pain reliever. The results of this study may very well have been different if another type of prescription drug had been used, because of participants' baseline beliefs about prescription painkillers in particular, as well as the fact that they are used to relieve symptoms (pain), as opposed to actually treating a chronic medical condition. For example, the same effects may not have been found if the message had been about the risk of side effects from a prescription anti-anxiety medication (benzodiazepines). This is due not only to the different uses for the drug, but also likely because of individuals' preconceptions and the media's (lack of) coverage.

One much less frequently hears of problems with addiction to benzodiazepines, even though abuse and addiction are problems with these medications, as well (Griffiths & Johnson, 2005). Although much is said about the risk of overdose from prescription painkillers, benzodiazepines present this danger as well (especially when mixed with alcohol.) Therefore, prescription pain medications may be emotionally charged in a way that is unique from other kinds of prescriptions, limiting the effects here from being generalized more widely. Indeed, the

67

emotional reaction to this kind of drug (and the experiential influences) may significantly impact how individuals interpret new information about the risks of prescription pain medications.

Since this study would hopefully be useful for risk perception situations outside of the processing of information about prescription drugs, it is also worth considering how these findings might be applied to different topics, such as environmental risk and climate change. The use of verbal quantifiers to describe risks from these issues may not be as emotionally charged or perceived as quite as relevant to individuals' personal experiences. This could have an impact on how individuals' interpret these messages and the effect they have on perceptions and behaviors. Even in other medical risk contexts, such as a patient-provider context, these effects may be different. For example, a physician may be seen as a more trusted source of health risk information than a pharmaceutical advertisement, since the latter is intended to sell products, instead of focusing on individuals' health. Additionally, a prescription drug ad is designed to appeal and be relevant to a wide variety of individuals, while a physician has an intimate and extensive knowledge of an individual's personal and medical history. These factors may help to explain why individuals might more drastically inflate their interpretations of terms such as "very rare" in this situation, compared with others.

Another possible limitation has to do with the age of the individuals in the sample used in the study. Although the lower limit of the age range was cut off at 25 partly because of the smaller likelihood of younger people having used prescription pain medications (for prescribed, and not recreational purposes, at least), the upper limit of 55 excludes a large number of relevant individuals. This is particularly true since the likelihood of using prescription drugs more generally tends to increase as individuals get older, with individuals who are 65 and older being the biggest consumers of prescription medications (Center for Substance Abuse Treatment,

68

2012). Researchers in future studies might be wise to include older populations of participants in their sample for this reason.

Another possible limitation is the fact that only two types of side effects were studied. In particular, they are side effects from rather extreme ends of the spectrum. Dry mouth may be viewed as such a small side effect that it hardly registers. On the other hand, "intense nausea and vomiting" may be seen as so severe that it could blunt participants' more nuanced interpretation of likelihood. Although the adjective "intense" was added to the high severity side effect to help ensure that participants actually saw it as quite severe, it is very possible that it was a bit heavyhanded. In other words, it may have been unrealistically extreme and participants may interpreted anything with the prefix of "intense" as being an unduly large risk to take in when it comes to a prescription drug.

It would also be valuable to have some conditions where participants were given the option of an open-ended response for the numerical risk estimate measure. This way, a comparison could be made between those who used the sliding scale and those who simply produced the numbers for an open-ended response. It is entirely possible that seeing a scale caused participants to naturally inflate the estimates that they gave, so that they would be higher than if they had simply typed them into a blank space on the survey. Since the main interest here was in how people would interpret these terms in relation to each other—and therefore an upwards bias in the measurement would have presumably affected all estimates indiscriminately—this was not considered to be a major concern in the study's design. Additionally, the participants' inflated estimates for each of these interpretations are largely consistent with the studies conducted by Dianne Berry, who used open-ended responses for

participants' estimates. Therefore, it would appear that the measurement type likely didn't have a huge effect on participants' responses.

APPENDIX

APPENDIX A

Risk Information Study Questionnaire

I1 CONSENT FORM

I2 Do you understand this information and consent to participate in this study?

- o Yes
- o No

I3 Are you between ages 25-55?

- o Yes
- o No

I5 I agree to carefully read all of the information in this survey and provide thoughtful and honest answers to all of the questions

- o Yes
- o No

I4 Have you ever taken a pain relief medication (prescription or over-the-counter)?

- o Yes
- o No

I6 Thanks for your willingness to participate. In a moment, you will have at least 15 seconds to review a description of a hypothetical prescription pain medication. Please read this description carefully. After the 15 seconds, the forward button will appear. You can click on this button to answer some questions about the prescription pain medication, Soulagis.

S1 Manufacturer's product description:

Prescription Soulagis caplets provide temporary relief of moderate-to-severe pain. A single caplet provides up to 12 hours of pain relief. Soulagis is intended for adults and children 12 years and older. Take with a full glass of water.

Very common side effects of this medication include dry mouth.

S2 Manufacturer's product description:

Prescription Soulagis caplets provide temporary relief of moderate-to-severe pain. A single caplet provides up to 12 hours of pain relief. Soulagis is intended for adults and children 12 years and older. Take with a full glass of water.

<u>Common</u> side effects of this medication include <u>dry mouth</u>.

S3 Manufacturer's product description:

Prescription Soulagis caplets provide temporary relief of moderate-to-severe pain. A single caplet provides up to 12 hours of pain relief. Soulagis is intended for adults and children 12 years and older. Take with a full glass of water.

Rare side effects of this medication include dry mouth.

S4 Manufacturer's product description:

Prescription Soulagis caplets provide temporary relief of moderate-to-severe pain. A single caplet provides up to 12 hours of pain relief. Soulagis is intended for adults and children 12 years and older. Take with a full glass of water.

Very rare side effects of this medication include dry mouth.

S5 Manufacturer's product description:

Prescription Soulagis caplets provide temporary relief of moderate-to-severe pain. A single caplet provides up to 12 hours of pain relief. Soulagis is intended for adults and children 12 years and older. Take with a full glass of water

Side effects of this medication include dry mouth.

S6 Manufacturer's product description:

Prescription Soulagis caplets provide temporary relief of moderate-to-severe pain. A single caplet provides up to 12 hours of pain relief. Soulagis is intended for adults and children 12 years and older. Take with a full glass of water.

Very common side effects of this medication include intense nausea and vomiting.

S7 Manufacturer's product description:

Prescription Soulagis caplets provide temporary relief of moderate-to-severe pain. A single caplet provides up to 12 hours of pain relief. Soulagis is intended for adults and children 12 years and older. Take with a full glass of water.

<u>Common</u> side effects of this medication include <u>intense nausea and vomiting</u>.

S8 Manufacturer's product description:

Prescription Soulagis caplets provide temporary relief of moderate-to-severe pain. A single caplet provides up to 12 hours of pain relief. Soulagis is intended for adults and children 12 years and older. Take with a full glass of water.

Rare side effects of this medication include intense nausea and vomiting.

S9 Manufacturer's product description:

Prescription Soulagis caplets provide temporary relief of moderate-to-severe pain. A single caplet provides up to 12 hours of pain relief. Soulagis is intended for adults and children 12 years and older. Take with a full glass of water.

Very rare side effects of this medication include intense nausea and vomiting.

S10 Manufacturer's product description:

Prescription Soulagis caplets provide temporary relief of moderate-to-severe pain. A single caplet provides up to 12 hours of pain relief. Soulagis is intended for adults and children 12 years and older. Take with a full glass of water.

Side effects of this medication include intense nausea and vomiting.

QR1 If 1,000 patients took Soulagis (the prescription you just read about), around **how many** of these patients would you expect to experience the listed side effect? (Please indicate the number using the slider below).

0		 	1000
Users affected by the	e side effect		

QR2 How confident are you in the estimate you gave above?

- Very confident
- Somewhat confident
- Neither confident nor unconfident
- Somewhat unconfident
- Very unconfident

QR3 Sometimes we are uncertain about how many times something will occur, so we indicate a **range** of likely values. For example, if you flipped a coin 100 times, you might predict that it is likely to come up "heads" somewhere between 40 and 60 times.

Similarly, in the question below, please indicate the **range** of users (out of 1,000) who are likely to experience the side effect after taking this prescription

If 1,000 people took this prescription, I would expect **between** _____ and _____ of these users to experience the listed side effect.

0	1000
Low estimate	1000
0	
High estimate	1000

QR4 How confident are you in the estimates you gave above?

- Very confident
- Somewhat confident
- Neither confident nor unconfident
- Somewhat unconfident
- Very unconfident

QR5 What were the main **risks** you thought about while reading this description? Please write as much as you like.

Q102 What were the main **benefits** you thought about while reading this description? Please write as much as you like.

Q108 The message you just read described a potential side effect of taking this drug.

Please rate how severe, serious, upsetting, or mild it would be if you experienced the side effect:

The side effect that was listed in the drug description was:

	1	2	3	4	5	6	7	8	9
Not severe: Severe	0	0	Ο	Ο	Ο	Ο	Ο	Ο	0
Not serious: Serious	0	0	0	0	О	0	О	Ο	0
Not upsetting: Upsetting	o	o	0	0	О	Ο	О	Ο	ο
Mild: Not mild	0	0	0	0	О	Ο	Ο	0	0

Q110 Please type the side effect that was listed in the description you just read:

Q110 What was the likelihood of getting the side effect in the description you read?

- No mention of likelihood
- Very rare
- o Rare
- Somewhat rare
- Somewhat common
- \circ Common
- Very common

Q83 The following questions will ask you about your intended behaviors regarding this medication.

I would ask my physician for more information about this medication:

- o Strongly agree
- o Agree
- Neither agree nor disagree
- o Disagree
- o Strongly disagree

Q84 I would search for more information about this medication online:

- o Strongly agree
- o Agree
- Neither agree nor disagree
- o Disagree
- o Strongly disagree

Q85 I would purchase or use this medication:

- Strongly agree
- o Agree
- Neither agree nor disagree
- o Disagree
- Strongly disagree

Q86 I would recommend this medication to a friend or family member:

- o Strongly agree
- o Agree
- Neither agree nor disagree
- o Disagree
- Strongly disagree

Q101 The potential side effects of this drug are very severe:

- Strongly agree
- o Agree
- Neither agree nor disagree
- o Disagree
- Strongly disagree

Q98 If I took this drug, I would be likely to experience side effects:

- Strongly agree
- o Agree
- Neither agree nor disagree
- o Disagree
- Strongly disagree

Q99 If I took this drug, I would be worried about side effects:

- o Strongly agree
- o Agree
- Neither agree nor disagree
- o Disagree
- Strongly disagree

Q105 The medication is effective at relieving pain:

- o Strongly agree
- o Agree
- Neither agree nor disagree
- o Disagree
- Strongly disagree

PP1 The following questions will ask you about your impressions of and experiences with prescription pain medications. Please indicate your level of agreement with the statements given.

Prescription pain medications have a lot of side effects:

- o Strongly agree
- o Agree
- Neither agree nor disagree
- o Disagree
- Strongly disagree

PP2 Prescription pain medications are dangerous to those who take them:

- Strongly agree
- o Agree
- Neither agree nor disagree
- o Disagree
- Strongly disagree

PP3 How likely do you think it is that a person taking a strong prescription pain medication will become addicted to it?

- Very likely
- Somewhat likely
- Neither likely nor unlikely
- Somewhat unlikely
- Very unlikely

PP6 The following questions ask for your perceptions of prescription pain medication abuse in your state.

Below, please select the state or territory in which you currently reside.

- o Alabama
- o Alaska
- o Arizona
- o Arkansas
- o California
- o Colorado
- Connecticut
- o Delaware
- o Florida
- o Georgia
- o Hawaii
- o Idaho
- o Illinois
- o Indiana
- o Iowa
- o Kansas
- o Kentucky
- o Louisiana
- o Maine
- o Maryland
- o Massachusetts
- o Michigan
- o Minnesota
- o Mississippi
- o Missouri
- o Montana
- o Nebraska
- o Nevada
- New Hampshire
- o New Jersey
- o New Mexico
- o New York
- o North Carolina
- North Dakota
- o Ohio
- o Oklahoma
- o Oregon
- o Pennsylvania
- Rhode Island

- o South Carolina
- o South Dakota
- o Tennessee
- o Texas
- o Utah
- o Vermont
- o Virginia
- o Washington
- West Virginia
- Wisconsin
- Wyoming
- District of Columbia (Washington D.C.)

PP7 How serious do you think the abuse of prescription pain medications is in your state?

- Extremely serious
- Very Serious
- Somewhat serious
- Not too serious
- Not a problem at all
- o Don't know

PP8 Do you believe that the problem of prescription pain medication abuse in your state is **better, worse**, or about **the same** as it was 5 years ago?

- o Better
- o Worse
- o Same
- o Don't know

PP9a For each of the following, please state whether or not you think it is a **major** cause, a **minor** cause, or **not** a cause of abuse of prescription pain medications in your state:

It is too easy to buy prescription painkillers illegally:

- o Major cause
- o Minor cause
- Not a cause
- o Don't know

PP9b It is too easy to get prescription pain medications from people who have saved them from their old prescriptions:

- o Major cause
- Minor cause
- Not a cause
- o Don't know

PP9c Pain medications are prescribed too often, or in doses greater than what is needed:

- o Major cause
- o Minor cause
- Not a cause
- o Don't know

PP10 During the past 5 years, have you known anyone who has abused prescription pain medications?

- o Yes
- o No
- o Don't know

Q103 How often have you seen news coverage about prescription pain medication abuse and addiction?

- o Very often
- Occasionally
- o Rarely
- Not at all

Q104 How often have you spoken with friends or family about prescription pain medication abuse and addiction?

- o Very often
- Occasionally
- o Rarely
- o Not at all

Q98 The following questions aim to gauge your comfort and preference for using numbers. For each of the following questions, please check the box that **best reflects** how good are you at doing the following things.

How good are you at working with fractions?

- Not good at all
- 0
- 0
- 0
- 0
- Extremely good

Q100 How good are you at working with percentages?

• Not good at all

0

- 0
- 0
- 0
- Extremely good

Q102 How good are you at calculating a 15% tip?

- Not good at all
- 0
- 0
- 0
- 0
- Extremely good

Q104 How good are you at figuring out how much a shirt will cost if it is 25% off?

- Not good at all
- 0
- 0
- 0
- 0
- Extremely good

Q108 For each of the following questions, please check the box that **best reflects** your answer:

When reading a newspaper, how **helpful** do you find tables and graphs that are parts of the story?

- Not helpful at all
- 0
- 0
- 0
- 0
- Extremely helpful

Q110 When people tell you the chance of something happening, do you prefer that they use **words** ("it rarely happens") or **numbers** ("there's a 1% chance")?

Always prefer words
Always prefer numbers

Q112 When you hear a weather forecast, do you prefer predictions using **percentages** (e.g., "there will be a 20% chance of rain today") or predictions using **words** (e.g., "there is a small chance of rain today")?

- o Always prefer percentages
- 0

0

0

0

• Always prefer words

Q114 How often do you find numerical information to be useful?

o Never

0

- 0
- 0

0

• Very often

Q118 The following questions will ask you to do a few different types of **mathematical calculations**. Please write your answer in the space provided:

Imagine that you rolled a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up even (rolling a 2,4, or 6)?

Q120 In the Big Bucks Lottery, the chance of winning a \$10.00 prize is 1%. What is your guess about how many people would win a \$10.00 prize if 1,000 people each buy a single ticket to the lottery?

Q122 In the ACME Publishing Sweepstakes, the chance of winning a car is 1 in 1,000. What percent of tickets to the ACME Publishing Sweepstakes win a car?

Q124 Which of the following numbers represents the biggest risk of getting a disease?

- o 1 in 100
- o 1 in 1,000
- o 1 in 10

Q126 Which of the following numbers represents the biggest risk of getting a disease?

- o 1%
- o 10%
- o 5%

Q128 If Person A's risk of getting a disease is 1% in ten years, and Person B's risk is double that of A's, what is B's risk?

Q130 If Person A's chance of getting a disease is 1 in 100 in ten years, and Person B's risk is double that of A's, what is B's risk?

Q132 If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100?

Q134 If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1,000?

Q136 If the chance of getting a disease is 20 out of 100 this would be the same as having a ____% chance of getting the disease?

Q138 The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected?

Q88 Now we would like to ask you some questions about you personally, to help us classify the data:

What is your gender?

- o Male
- o Female
- Transgender
- No answer

Q89 Into which category does your current age fall?

- 25-29 years old
- o 30-34 years old
- o 35-39 years old
- 40-44 years old
- 45-49 years old
- o 50-55 years old

Q90 What is the highest level of education you have completed and received credit for?

- Some high school
- High school graduate
- Some college
- o College graduate
- Some post graduate work
- Post graduate degree
- Vocational

Q91 Which of the following categories best captures your ethnicity?

- Asian-American/Pacific Islander
- African-American/Black
- o Caucasian/White
- Hispanic/Latino
- Native American
- o Multiracial
- Other

Q92 Are you currently employed for pay?

- o Yes
- o No

Q93 What was your total household income last year before taxes?

- o Less than \$10,000
- o \$10,000-\$19,999
- o \$20,000 \$29,999
- o \$30,000 \$39,999
- o \$40,000 \$49,999
- o \$50,000 \$59,999
- o \$60,000 **-** \$69,999
- o \$70,000 \$79,999
- o **\$80,000 \$89,999**
- o **\$90,000 \$99,999**
- o **\$100,000 \$149,999**
- More than \$150,000

REFERENCES

REFERENCES

- Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50, 179-211. doi:10.1016/0749-5978(91)90020-t
- American Society of Addiction Medicine. (2016). Opioid addiction 2016 facts and figures. Retrieved from <u>www.asam.org/docs/default-source/advocacy/opioid-addiction-</u> <u>disease-facts-figures.pdf</u>
- Avorn, J., & Shrank, W. H. (2009). Communicating drug benefits and risks effectively: There must be a better way. *Annals of Internal Medicine*, *150*, 593-596. doi:10.7326/0003-4819-150-8-200904210-000012
- Berry, D., Michas, I., & Bersellini, E. (2002). Communicating information about medication side effects: Effects on satisfaction, perceived risk to health, and intention to comply. *Psychology & Health*, 17, 247-267. doi:10.1080/08870440290029520a
- Berry, D. C., Knapp, P., & Raynor, D. K. (2002). Provision of information about drug side-effects to patients. *The Lancet*, *359*, 853-854. doi:10.1016/s0140-6736(02)07923-0
- Berry, D. C., Raynor, D. K., & Knapp, P. (2003). Communicating risk of medication side effects: An empirical evaluation of EU recommended terminology. *Psychology, Health & Medicine*, 8, 251-263. doi:10.1080/1354850031000135704
- Berry, D., Raynor, T., Knapp, P., & Bersellini, E. (2004). Over the counter medicines and the need for immediate action: A further evaluation of European Commission recommended wordings for communicating risk. *Patient Education and Counseling*, 53, 129-134. doi: 10.1016/s0738-3991(03)00111-3
- Bonnefon, J. F., & Villejoubert, G. (2006). Tactful or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science*, 17(19), 747-751. doi:10.1111/j.1467-9280.2006.01776.x
- Boston Globe & Harvard School of Public Health. (2015). Prescription painkiller abuse: Attitudes among adults in Massachusetts and the United States. <u>https://cdn1.sph.harvard.edu/wp-content/uploads/sites/21/2015/05</u> /Prescription-Painkiller-Poll-Report.pdf
- Brainerd, C. J., & Kingma, J. (1984). Do children have to remember to reason? A fuzzy-trace theory of transitivity development. *Developmental Review*, *4*, 311-377. doi:10.1016/0273-2297(84)90021-2
- Brainerd, C. J., & Kingma, J. (1985). On the interdependence of short-term memory and working memory in cognitive development. *Cognitive Psychology*, 17, 210-247. doi:10.1016/0010-0285(85)90008-8

- Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new institutionalism. *Developmental Review*, *10*, 3-47. doi:10.1016/0273-2297(90)90003-m
- CDC. (2013) Addressing prescription drug abuse in the United State: Current Activities and Future Opportunities. https://www.cdc.gov/drugoverdose/pdf/hhs_prescription_drug_abuse_report_09.2013.pdf
- CDC. (2015). Health, United States, 2015. https:// https://www.cdc.gov/nchs/hus/index.htm
- Cliff, N. (1959). Adverbs as multipliers. *Psychological Review*, 66(1), 27-44. doi:10.1037/h0045660
- Cliff, N. (1972). Adverbs multiply adjectives. In J. M. Tanur (Ed.), Statistics: A Guide to the unknown (pp. 176-184).
- Cacioppo, J. T., von Hippel, W., Ernst, J. M. (1997). Mapping cognitive structures and processes through verbal content: The thought-listing technique. *Journal of Consulting and Clinical Psychology*, 65, 928-940. doi:10.1037//0022-006x.65.6.928
- Center for Substance Abuse Treatment. (2012). *Substance Abuse Among Older Adults*. Treatment Improvement Protocol (TIP) Series, No. 26. HHS Publication No. (SMA) 12-3918. Rockville, MD: Substance Abuse and Mental Health Services Administration.
- Chaiken, S. (1980). Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion. *Journal of Personality & Social Psychology*, 39(5), 752-766. doi:10.1037//0022-3514.39.5.752
- Cox, J. (2016). Verbal quantifiers, numeracy, and enhancing risk communication. *Unpublished paper*.
- DHHS. (2006). *Guidance for industry: Adverse reactions section of labeling for human* prescription drug and biological products--content and format. Rockville, MD: US
- DHHS. (2013). Highlights of the 2011 Drug Abuse Warning Network (DAWN) findings on drug-related emergency department visits. Rockville, MD: US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration http://www.samhsa.gov/data/2k13/DAWN127/sr127-DAWN-highlights.htm
- European Commission. (1995). Frequency of adverse drug reactions. Guideline for preparing core clinical safety information on drugs. Reported from CIOMS working group III, Geneva, 1995.

- European Commission, Enterprise and Industry Directorate-General--Pharmaceuticals. (2009). A guideline on summary of product characteristics. Retrieved from http://ec.europa.eu/health/files/eudralex/vol-2/c/smpc_guideline_rev2_en.pdf
- Fagerlin, A., Zikmund-Fisher, B.J., Ubel, P.A., Jankovic, A., Derry, H.A., & Smith, D.M. (2007). Measuring numeracy without a math test: Development of the Subjective Numeracy Scale (SNS). *Medical Decision Making*, 27, 672-680. doi:10.1177/0272989x07304449
- Fagerlin, A., & Peters, E. (2011). Quantitative information. In B. Fischhoff, N. T. Brewer, & J. S. Downs (Eds.), *Communicating risks and benefits: An evidence-based user's guide* (pp. 53-64). Silver Spring, MD: US Department of Health and Public Services.
- FDA. (2016). Think it through: A guide to managing the risks and benefits of medicines. https:// https://www.fda.gov/downloads/drugs/resourcesforyou/ucm163235.pdf
- FDA. (2017) Understanding the influence of prescription drug advertising. https://www.fda.gov/Drugs/NewsEvents/ucm543439.htm
- Ghazal, S., Cokely, E. T., & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and Decision Making*, 9, 15-34. doi:10.1037/e573552014-017
- Gigerenzer G., Gaissmaier W., Kurz-Milcke E., Schwartz, L.M., & Woloshin, S. (2008). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8(2), 53-96. doi:10.1111/j.1539-6053.2008.00033.x
- Griffiths, R. R.; Johnson, M. W. (2005). Relative abuse liability of hypnotic drugs: A conceptual framework and algorithm for differentiating among compounds. *Journal of Clinical Psychiatry*, *66*, 31-41.
- Hartley, J., Trueman, M., & Rodgers, A. (1984). The effects of verbal and numerical quantifiers on questionnaire responses. *Applied Ergonomics*, *15*(2), 149-155. doi:10.1016/0003-6870(84)90332-6
- International Association for the Evaluation of Education Achievement. (2009). Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades. <u>https://timss.bc.edu/timss2007/PDF/TIMSS2007_InternationalMathematicsReport.pdf</u>
- IES. *National assessment of adult literacy* [Report]. (2003). Washington, DC, US: Institute of Education Sciences.
- Jones, C. M., Mack, K. A., Paulozzi, L. J. (2013). Pharmaceutical overdose deaths, United States, 2010. *JAMA*. 309(7), 657-9. doi:10.1001/jama.2013.272

Kahneman, D. (2011). Thinking, fast and slow. New York, NY: Farrar, Straus and Giroux.

- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 42(2), 263-291. doi:10.2307/1914185
- Kantor, E. D., Rehm, C. D., Haas, J. S., Chan, A. T., & Giovannucci, E. L. (2015). Trends in prescription drug use among adults in the United States 1999-2012. JAMA. 314(17), 1818-30. doi:10.1001/jama.2015.13766
- Kaphingst, K. A., DeJong, W., Rudd, R. E., & Daltroy, L. H. (2010). A content analysis of direct-to-consumer television prescription drug advertisements. *Journal of Health Communication*, 6, 515-528. doi:10.1080/10810730490882586
- Knapp, P., Raynor, D. K., & Berry, D. C. (2004). Comparison of two methods of presenting risk information to patients about the side effects of medicines. *Quality and Safety in Health Care*, 13, 176-180. doi:10.1136/qshc.2003.009076
- Kruger, J., & Dunning, D. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-34. doi:10.1037//0022-3514.77.6.1121
- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21(1), 37-44. doi:10.1177/0272989x010210010
- Loewenstein, G. G., Weber, E. U., Hsee, C. K., & Welch, N. (2001) Risk as feelings. *Psychologial Bulletin*, 127(2), 267-86. doi:10.1037//0033-2909.127.2.267
- Nelson, D. E., Hesse, B. W., & Croyle, R. T. (2009). *Making data talk*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780195381535.003.0004
- New York Times. (2017). Prescription drug abuse (archive). https://www.nytimes.com/topic/subject/prescription-drug-abuse
- Newstead, S. E., & Collis, J. M. (1987). Context and the interpretation of quantifiers of frequency. *Ergonomics*, *30*(10), 1447-1462. doi:10.1080/00140138708966038
- Okamoto, M., Kyutoku, Y., Sawada, M., Clowney, L., Watanabe, E., Ippeita, D., & Kawamoto, K. (2012). Health numeracy in japan: Measures of basic numeracy account for framing bias in a highly numerate population. *BMC Medical Informatics & Decision Making*, *12*. doi:10.1186/1472-6947-12-10

- Palos, G. R., Mendoza, T. R., Cantor, S. B., Aday, L. A., & Cleeland, C. S. (2004). Perceptions of analgesic use and side effects: What the public values in pain management. *Journal of Pain and Symptom Management*, 28(5), 460-473. doi:10.1016/j.jpainsymman.2004.02.016
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology: 129*, 123-205 doi:10.1016/s0065-2601(08)60214-2
- Pew Research Center. (2017). "U.S. students' academic achievement still lags that of their peers in many other countries." Pew Research Center, Washington, D.C.
- Reyna, V. F. (2008). A theory of medical decision making and health: Fuzzy trace theory. *Medical Decision Making*, *28*(6), 850-865. doi:<u>10.1177/0272989x08327066</u>
- Reyna, V. (2012). A new institutionalism: Meaning, memory, and development in fuzzy trace theory. *Judgment and Decision Making*, *7*, 332-359.
- Reyna, V. F., & Adam, M. B. (2003). Fuzzy-trace theory, risk communication, and product labeling in sexually transmitted diseases. *Risk Analysis*, 32(2), 325-342. doi:10.1111/1539-6924.00332
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. Learning and Individual Differences, 1(7), 1-75. doi: 10.1016/1041-6080(95)90031-4
- Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, 18(1), 89-107. doi:10.1016/j.lindif.2007.03.011
- Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, 135, 943-973. doi: <u>10.1037/a0017327</u>
- Ryan, C. L., & Bauman, K. (2016). Educational attainment in the United States: 2015. <u>https://www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf</u>
- Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annuls of Internal Medicine*, *127*, 966-972. doi:10.7326/0003-4819-127-11-199712010-0000 doi: <u>10.7326/0003-4819-127-11-199712010-00003</u>
- Slovic, P. (1987). Perception of risk. *Science*, *236*, 280.285. doi:<u>10.1126/science.3563507</u>

- Wallsten, T. S., Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language*, 25, 571-587. doi:10.1016/0749-596x(86)90012-4
- Zikmund-Fisher, B. J., Smith, D. M., Ubel, P. A., & Fagerlin, A. (2007). Validation of the subjective numeracy scale (SNS): Effects of low numeracy on comprehension of risk communications and utility elicitations. *Medical Decision Making*, *27*(5), 663-671. doi:10.1177/0272989x07303824