THE LANGUAGE OF RISK

A QUANTIATIVE STUDY ON THE EFFECTS OF PRESENTATION FORMAT IN RISK COMMUNICATION

ANDREAS LIDIN

DANIEL MIRA

Bachelor Thesis

Stockholm School of Economics

2023

The language of risk: A quantitative study on the effects of presentation format in risk communication

Abstract:

Consumers are faced with everyday decisions requiring understanding and interpretation of risks and probabilities. One area where the importance of good risk communication is particularly salient is in modern healthcare where shared medical decision making is a central feature. This thesis aimed to examine how presentation format might influence important outcome variables related to understanding, consumer satisfaction and risk perception. We took an experimental approach and used a quantitative method to conduct a between-participant study. A scenario experiment was conducted by distributing an online questionnaire with 152 participants. Subjects were randomly assigned to one of three treatment conditions (percentage, frequency, and visual) and were subsequently presented with a medical scenario. Findings indicate no significant effect on consumer satisfaction (measured through service performance and word-of-mouth intention), but in line with previous research there were significant differences between groups in terms of understanding. Risk perception was also significantly influenced by the presentation format when comparing relative risk. Results reveal that all three presentation formats are suitable when communicating basic and absolute risk, but frequency format seems most appropriate when communicating risks that require more cognitive processing.

Keywords:

Risk Communication, Numeracy, Icon arrays, Presentation Format, Natural Frequencies

Authors:

Andreas Lidin (24186)
Daniel Mira (24785)

Tutors:

Patric Andersson, Associate Professor, Department of Marketing and Strategy

Examiner:

Henrik Glimstedt, Associate Professor, Department of Marketing and Strategy

Bachelor Thesis
Bachelor Program in Business and Economics
Stockholm School of Economics

Contents

Acknowledgements

We would like to thank our supervisor, **Patric Andersson**, for his exceptional guidance and support. Your feedback has been essential throughout this project. For making time for meetings at short notice, answering e-mails any day of the week, and engaging in discussions with enthusiasm – thank you.

We would also like to express our gratitude to **friends and families** for constant encouragement and understanding.

To **all the respondents** in our questionnaire **–** your help has been invaluable.

Thank you to everyone who helped us make this thesis possible.

Andreas Lidin
Daniel Mira

1. Introduction

In today's society, consumers are expected to interpret and evaluate a myriad of statistical information in order to make informed decisions. These decisions range from the day-to-day questions of whether or not to wear a raincoat based on the weather forecast, to the more serious subjects of financial investments and health. Given these expectations, it might be surprising to learn from research that people tend to not always integrate probabilistic information in an adequate way (Hertwig et al., 2008). A variety of factors have the potential to impact or bias an individual's understanding of risk and subsequent behavior.

One area where the importance of good risk communication is particularly salient is in modern health care. Patients are expected to assimilate numerical information about risks associated with certain procedures and medications (or lack thereof) and based on this make informed decisions that lie in their own best interest. Failure to do so can lead to adverse side effects at best, or even death at worse. In light of the recent pandemic the importance of good risk communication became even more palpable as governments all over the world tried to contain the spread of the virus and drive herd immunity through vaccination programmes.

This thesis will examine how different representations of risk can influence important outcome variables related to understanding, consumer satisfaction and risk perceptions. We chose to limit ourselves to the medical domain as scenarios pertaining to this lie close to most individuals and allow for realistic representations of risk.

1.1. Background

1.1.1. Factors influencing risk perception

A variety of factors have the potential to impact people's perception of risk. One of the more well-known examples comes from the framing effect in which individuals tend to be more risk-averse when decision outcomes are framed in terms of gains, and risk-seeking in conditions of losses (Kahneman & Tversky, 1984).

It is not just framing or contextual differences that have the ability to influence people. Sometimes even correctly understanding the risk in itself can be hard. Without an adequate understanding of the fundamental premises of a decision – no one can be expected to make rational choices. Research has shown that the presentation format of statistical information can significantly influence individuals' ability to correctly assess and integrate probabilistic information (McDowell & Jacobs, 2017). More specifically, in a meta-analysis spanning 20 years of research and 35 articles the authors found that presenting probabilities as frequencies improved participants' statistical reasoning and by including visual aids this effect was even stronger (McDowell & Jacobs, 2017).

1.1.2. Individual characteristics and proposed solutions

Another important finding is the interaction of an individual's numeracy and ability to correctly interpret probabilities, where numeracy seems to predict to what degree an individual is susceptible to misinterpretations (Galesic et al., 2009). Importantly, it has been shown that statistically naive individuals can significantly improve their statistical reasoning by altering the numerical presentation format from percentages to frequencies (Gigerenzer & Hoffrage, 1995). The mechanism that this works through has been argued to be that frequency formats more closely mimics the "sequential way information is acquired in natural sampling" (Gigerenzer & Hoffrage, 1995). Much as a pocket calculator cannot receive written information as input format to perform complex computations, the human mind is not evolved to receive abstract statistical point estimates in order to make predictions about the real world (Gigerenzer & Hoffrage, 1995). One fundamental solution to this problem is to simply communicate probabilities as frequencies (McDowell & Jacobs, 2017; Gigerenzer & Hoffrage, 1995). Taking it one step further would be to include visual aids (specifically icon arrays), which has shown promising results in increasing accuracy and understanding of medical risks for both high- and low-numeracy individuals (Galesic et al., 2009).

1.1.3. Shared medical decision making

Shared medical decision making is a central feature in modern healthcare. Taking Sweden as an example, the patient is viewed as a co-producer of the service and the communication with clinicians is a pivotal part of the person-centered care (Hedberg et al., 2022). Clinicians must therefore make careful attempts to effectively communicate estimated risks and benefits of various treatments. Unless patients have a thorough understanding of these fundamental premises, they will not be able to make informed decisions about their own health (Stovitz & Shrier, 2013).

1.2. Problem area

The importance of understanding statistical information in today's society highlights the need to examine how risk communication affects consumers. This is especially true in the domain of modern healthcare where shared medical decision making is a central feature. Although many previous studies have investigated the effects of presentation format on understanding in a test environment, we wanted to insert these findings into a realistic medical scenario. Against this background we also want to examine how presentation format can influence outcome variables related to consumer satisfaction and risk perception. To the best of our knowledge, little to none previous research has extended their analysis to a scenario experiment and included more adjacent variables associated with risk communication.

This is relevant to marketing in several ways. One important area is consumer behavior and decision-making. Understanding how consumers interpret risk can give valuable insights into how they then process this information and form perceptions which informs the choices they make. The healthcare industry needs to communicate numerical information to consumers on an ongoing basis. Any findings related to this problem area can be of importance for that industry or any actor involved in communicating risk.

1.3. Research purpose and questions

This thesis aims to address questions related to communication and representation of risk. The purpose is to understand how consumers are affected by presentation format in a general sense, and how this should shape risk communication in order to support them. The first objective is to simply ascertain whether presentation format actually has a significant effect on participants' understanding of statistical information. The next objective is to try to understand how adjacent variables related to risk communication is affected. More concretely, does presentation format influence variables related to consumer satisfaction? A third objective is to investigate if and how presentation format might influence absolute and relative risk perceptions.

1.4. Delimitations

This thesis is subject to a number of delimitations. First of all, we chose to only examine the effects of presentation format in the medical domain as time restraints made it impractical to consider other contexts. Secondly, the geographical scope is limited to Sweden and Swedish-speaking participants. We also had limited resources which means that we had to use convenience sampling for data collection. For this reason, the generalizability of our findings might not extend beyond our sample group.

We also had to adapt to GDPR and ethical considerations as our chosen subject was associated with health care. We made careful attempts as to not collect any sensitive data related to the respondents' personal health status and made sure that the medical scenario was always referring to a third person. Careful considerations were made on what demographic variables should be collected and we only included the following for analysis: age, gender, and educational attainment. All data was collected and handled in accordance with GDPR guidelines.

1.5. Expected contribution

With this thesis we hope to contribute to the existing body of research on consumer behavior and decision-making in general and risk communication in particular. Through our experimental design using a medical scenario we aspire to capture new insights

related to consumer satisfaction and risk perception following exposure to different presentation formats. By applying existing research findings in a scenario, we hope to contribute with practical implications that can extend beyond the traditional test environment.

1.6. Thesis outline

In an attempt to answer our research questions this thesis is divided into five different sections: 1) Introduction, 2) Literature review, 3) Methodology, 4) Results and analysis, and 5) Discussion. Following this introduction, the next section will cover previous research and the theoretical foundations for our study. After that we will describe the scientific approach that was chosen. The results and analysis section will include a presentation of the collected data and the outcomes of our statistical analysis. In the discussion we will synthesize our results and discuss the implications and limitations of our study.

2. Literature review

2.1. Literature review approach

In order to find relevant research and articles on the subject of risk communication and presentation format we conducted a systematic search in the most common library databases. The search was conducted by including keywords such as: *risk communication, *numeracy, *presentation format, *icon arrays, and *natural frequencies.

2.2. Risk communication

Risk communication involves theories from both the natural and social sciences, where studies of human perception and decision making are especially relevant. It can be defined as communication in general or as the "process by which information is transferred by one party to another through a variety of channels". The effectiveness of this communication can be defined both as the degree to which the message recipient understands the information or – as an extension – to what degree this understanding results in behavioral changes (Kamrin, 2014). The health belief model posits that people engage in self-protective actions when they believe that they are vulnerable to risk, that their actions will reduce this vulnerability and that the benefits of doings so is lower than the costs (DiClemente & Jackson, 2017). Before an individual can form a belief about their vulnerability however, they need to have an accurate understanding of the risk that they are exposed to. There are a variety of factors that can affect the success of risk communication, ranging from the trust in the source of the information to the form and content of the message. One critical part is the message recipient's perception which depends on how people internalize information in general, but also how that interacts with individual characteristics (Kamrin, 2014).

2.3. Previous studies

2.3.1. Natural frequencies

Perception and understanding of risk have been studied extensively and one important factor is the presentation format of statistical information. McDowell & Jacobs (2017) conducted a large-scale meta-analysis of the effect of natural frequencies on statistical reasoning. The authors found that when information was presented as frequencies instead of percentages, participants performed better when solving statistical inference problems. This positive effect was consistent even when they accounted for a host of potential moderators, spanning from individual characteristics like numeracy to problem representations (McDowell & Jacobs, 2017). These findings are supported by

Gigerenzer & Hoffrage (1995) who also found that individuals performed significantly better when presented with frequencies. They argued that frequency format is more comparable to how information is naturally gathered in the real world which makes inferential tasks (like Bayesian calculations) computationally easier to perform. They explained that throughout evolution, humans have collected information through series of events rather than by receiving abstract single point estimates such as percentages. This sequential collection of information resembles what is referred to as "natural sampling". For this reason, humans are more adapt at handling natural frequencies when making predictions about the real world (Gigerenzer & Hoffrage, 1995).

Another reason natural frequencies seem to make inferential tasks easier is because they include information about base rates. This feature of natural frequencies means that they provide information about how often an event occurs within a population without any conditional information. The result is that less calculations have to be performed in order to arrive at an answer when making statistical inferences (Hoffrage et al., 2000).

Furthermore, it is not just that expressing statistics as natural frequencies improves statistical reasoning, but probabilities (expressed as percentages) actually impedes people's abilities. In an experiment, when given percentages, only 1out of 24 physicians were able to correctly determine the number of patients with a positive screening test that truly had cancer. When another group of physicians were provided with frequencies, this number increased to 16 out of 24. The positive effect of frequency format has shown to apply to both experts and everyday people (Hoffrage et al., 2000).

2.3.2. Numeracy

A particularly important area of risk communication within health care is numeracy. Within the wider concept of health literacy, numeracy is one component. It refers to an individual's ability to handle numerical concepts and comprises basic math skills (Apter et al., 2008). Having low numeracy has been linked to adverse health outcomes for patients and a lower likelihood to follow recommendations by health care providers (Ruiz et al., 2013). In a study on U.S. literacy, it was found that 26% of adults were severely limited in their quantitative abilities and 32% had only rudimentary skills. In a similar vein, when investigating numeracy in a health care setting, many patients were unable to understand basic medical instructions. Almost one third were incapable of determining how many pills they should take based on their prescription (Rothman et al., 2008). Based on the importance of numeracy, many studies have set out to find ways to overcome it, with frequencies and visual aids being two proposed solutions (Galesic et al., 2009).

2.3.3. Visual format

Spiegelhalter et al. (2011) has emphasized the benefits of using graphics and visualization of data to communicate risks and uncertainties. The author pointed out how communicating probabilities to the general public can be facilitated by graphic visualization but also how the relative numeracy of the audience members is of importance (Spiegelhalter et al., 2011). One way to make the data more accessible is by using icon arrays. By displaying a number of figures or icons that represent individuals affected by a risk while simultaneously showing those not affected, the audience is better equipped to assimilate the information (Galesic et al., 2009). This visual aid is becoming more popular although previous research emphasize that experimental studies on its efficacy is still lacking (Galesic et al., 2009).

One of the more promising uses of graphical visualization is in enhancing understanding for individuals with low numeracy (Garcia-Retamero & Dhami, 2011). Particularly icon arrays have shown positive results in explaining relative risk changes to this group. A critique against its use is that in some cases it has been shown that including icon arrays can increase the perceived seriousness of risk. Since its use draws attention to the level of individuals (when figures are used) it has been argued that respondents might perceive the risks more vividly which adds affective components to the equation (Galesic et al., 2009).

2.3.4. Icon arrays

Previous studies on icon arrays have shown conflicting evidence on its efficacy. Ruiz et al. (2013) found that for male veterans in the U.S., risk understanding was lower with icon arrays and no differences were found for recall. However, Galesic et al. (2009) found contrary evidence indicating that the inclusion of icon arrays increased accuracy of understanding for both high- and low-numeracy individuals. There has also been empirical support indicating that using icon arrays improves gist and verbatim knowledge (Recchia et al., 2022). One study showed that using icon arrays in combination with numerical descriptions increased participants ability to estimate relative risk reductions and this effect was particularly strong for individuals with low numeracy. Another finding was that using a large number of icons made the risk appear more serious. This was explained by the ratio-bias-effect where a larger denominator inflates the perceived risk (Galesic et al., 2009). Despite inconclusive findings, official guidelines from government agencies (like the U.S. Department of Health and Human Services) advocate the use of pictographs and visual aids (Fischhoff, 2011).

2.4. Hypotheses and exploratory approach

Based on previous studies, two guiding hypotheses have been derived related to understanding. For all other variables, an exploratory approach will be used. Although conflicting evidence has been found on the efficacy of visual format (in particular of icon arrays), a consensus seems to exist on using pictographs and visual aids when possible (Fischhoff, 2011; Spiegelhalter et al., 2011). For this reason, a preliminary hypothesis for visual format is included.

> **H1:** Individuals exposed to frequency format will have better understanding compared to individuals exposed to percentage format.
>
> **H2:** Individuals exposed to visual format will have better understanding compared to individuals exposed to percentage format.

It is worth emphasizing that this study includes other outcome variables not directly related to understanding (although important and indirectly associated). For this reason, a large part of the thesis is devoted to the exploratory part. Some of our variables have only received peripheral attention in previous studies with inconclusive findings. Some have received no attention at all. Against this background, we view the hypothesis testing as secondary to our exploratory approach. Refer to Section 3.4. below for a complete overview of all included variables.

3. Methodology

3.1. Scientific approach

This thesis takes an experimental approach to understand the effects of presentation format. Moreover, we use a quantitative method in an attempt to capture between-participant effects by randomly assigning subjects to three different treatment conditions. The data is cross-sectional and we assume that the phenomenon is objective and external to observers, following an objectivist ontology (Bell et al., 2019). Furthermore, we will employ both inductive and deductive reasoning. In order to test existing theories, deduction will be used for the basic hypothesis testing. However, since we will also take an exploratory position, inductive reasoning will be used to see if our observations can help us draw any other conclusions (Bell et al., 2019).

Arguably, a qualitative approach could also prove useful in an attempt to gather deeper insights about the studied phenomenon (Bell et al., 2019). This method was however not chosen since the vast majority of previous studies on the subject employs a quantitative approach and we wanted to build on these findings. In addition, to conduct our experiment with different treatment effects, a self-completion questionnaire with randomized groups was considered to be the most appropriate. There are however obvious limitations to our chosen scientific method, one of the most prominent being the sampling strategy. Since we used convenience sampling, the generalizability of our findings is limited and the sample should not be viewed as representative of the population (Bell et al., 2019). As a result of resource and time constraints, this was however considered the most viable approach.

3.2. Pilot study

Before publishing the survey, we took precautions as to pre-test the questions and the scenario in its entirety. Using a pilot study group of ten people we had them take the first draft of the survey and asked for specific feedback on unclear questions, any formatting issues or other thoughts that came up. After this, we sat down with five of them and went over each block while having participants verbally comment on their thought process while answering the questions. We asked specifically about the relevance and formulation of the questions and their understanding of the material.

3.2.1. Pilot study insights

The pilot study group made important contributions to the final version of the questionnaire in their comments and by the way they interpreted some of the questions. Besides minor spelling mistakes that were brought up, many participants made us aware that the questionnaire took a long time to complete. We were aware of the phenomenon

of respondent fatigue and tried to address this by including a progress bar and provided clear instructions throughout the questionnaire on their progress and what remained. Some participants also mentioned that the survey felt like an exam which added unnecessary elements of anxiety and stress. For this reason, we added information about the fact that most of the questions did not have any "right" or "wrong" answers, but rather intended to gauge their perception or "gut-level assessment". Furthermore, one of the questions asked the respondents to estimate the percentage increase between two groups and provide an open entry answer. Considering the length of the survey and the aforementioned risk of survey fatigue, we choose to switch to a multiple-choice format for this particular question. Since participant dropout is directly linked to survey length, and because we prioritized a lower dropout rate above precise numerical estimates for this particular question, we choose to make it more convenient for respondents (Hoerger, 2010). All feedback was then incorporated into the final, revised version of the questionnaire.

3.3. Main study

3.3.1. Questionnaire

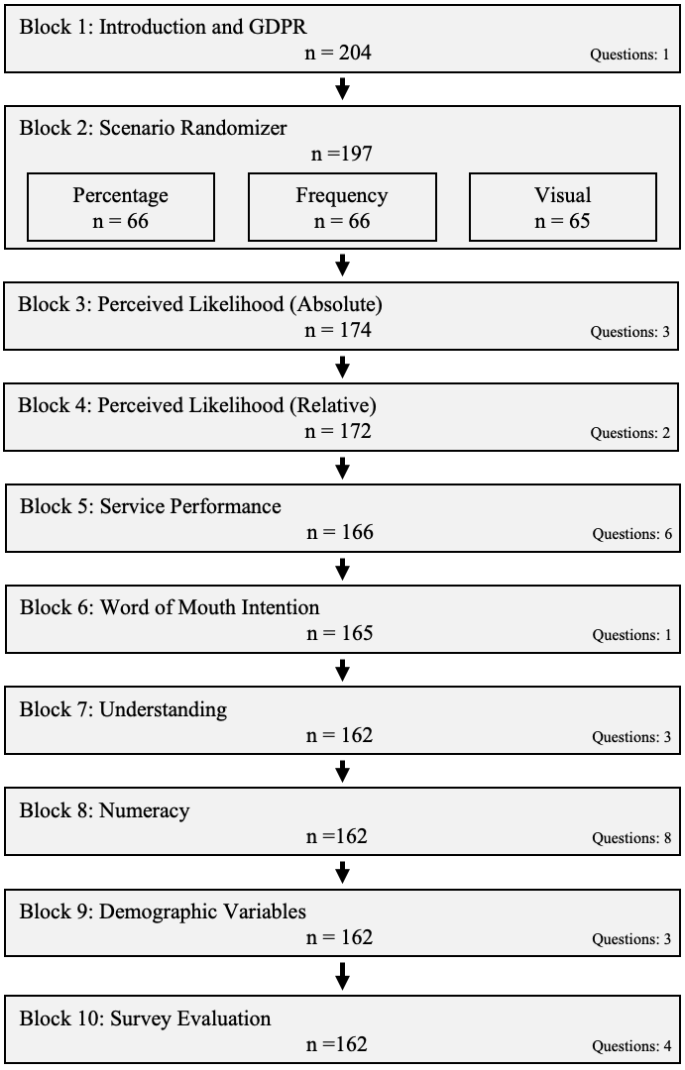
The study was carried out in the form of a self-completion questionnaire (see Appendix 7) which was constructed and administered through Qualtrics' online survey tool. We used the platform's functionality for anonymized responses which allowed us to adhere to GDPR guidelines. In total, the survey comprised of ten blocks: Block 1: Introduction and GDPR, Block 2: Scenarios, Block 3: Perceived Likelihood (Absolute), Block 4: Perceived Likelihood (Relative), Block 5: Service Performance, Block 6: Word-of-Mouth Intention, Block 7: Understanding, Block 8: Numeracy, Block 9: Demographic Variables, Block 10: Survey Evaluation.

Respondents started with the introduction where we gave a brief description of the aim of the study. We payed careful attention as to not reveal too much information since that could potentially influence the outcome of the experiment. Participants were then informed about the estimated time required to compete the survey, where to contact the authors and that 5 SEK would be donated to the Children's Cancer Fund (Barncancerfonden) for every submitted answer. Respondents were then asked to read the information about confidentiality and GDPR including the Consent Form for Participation Template. They were then asked to give their consent to participate by ticking a box and signing with their initials and today's date. This consent question had a force response functionalty and participants who ticked the "I do not consent" box were automatically sent out of the survey.

The next block randomly assigned each participant to one of three scenarios: one where they were exposed to statistical information presented in percentage format, the other in frequency format, and the last in a visual format. After this, participants answered

questions relating to the perceived likelihood of the negative event happening both from an absolute and from a relative (comparative) perspective. Respondents where then asked to rate the service performance based on the scenario they just read and then answer a question relating to their word-of-mouth intention. The next block included questions designed to test their understanding of the statistical information but would also serve to check the construct validity of our numeracy variable. Based on which presentation format participants had been randomly assigned to, we used branch logic in Qualtrics such that each group was once again exposed to the same risk information before they made their estimations. To measure numeracy, we used the Subjective Numeracy Scale (see section 3.4.3.). We then asked for some demographic information before having participants fill out a survey evaluation. Lastly we included an attention check question which asked the respondents what the scenario was about. We also included an instructional manipulation check (IMC) in block 5, as suggested by Oppenheimer et al. (2009).

3.3.2. Survey flow illustration



3.3.3. Experimental design

In order to test the effects of presentation format we chose to employ a scenario-based experiment in written form. To ensure that we had a representative design, we used an article from the Swedish publication Läkartidningen (the Physician Journal) about elevated blood pressure among young adults being a risk factor for future cardiovascular disease. In the article, the authors presented recent findings that 14 percent of young adults with normal blood pressure is expected to develop cardiovascular disease in the future, whereas 24 percent is expected for individuals with elevated blood pressure (Rietz & Brunström, 2023). We wanted to present these statistics in a realistic medical scenario and used one of the author's expertise (being a physician himself) to set up a representative description of a routine physical examination. To achieve the between-subject design we randomly assigned each participant to one of three treatment conditions: percentage format, frequency format and visual format. Every group was exposed to the same physical examination scenario with the only difference being which presentation format the risk statistics had. Participants received the information that the individual in the scenario had elevated blood pressure. See below for translated English versions of the three variations of the risk information:

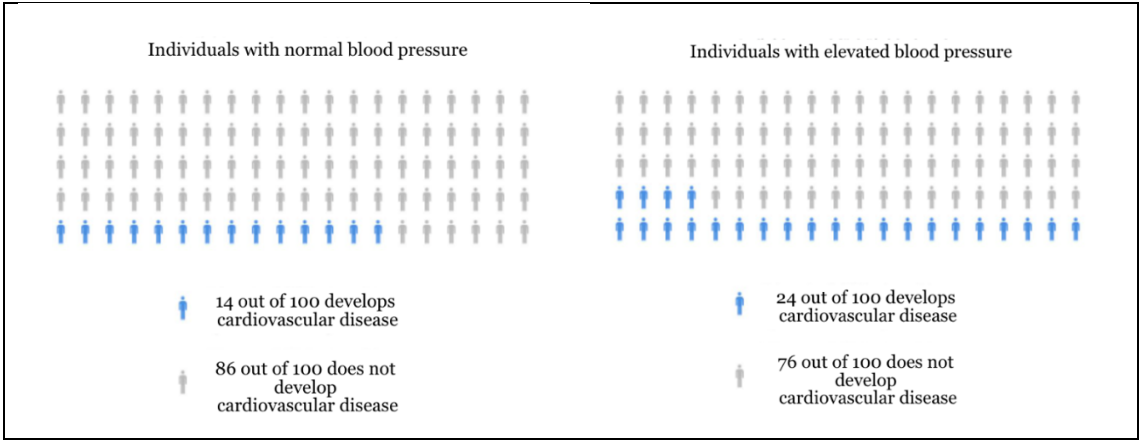
Percentage format

The physician further explains that for people with normal blood pressure, on average, **14%** will develop cardiovascular diseases in the future, whereas for people with elevated blood pressure, approximately **24%** will be affected.

Frequency format

The physician further explains that for people with normal blood pressure, on average, **14 out of 100** will develop cardiovascular diseases in the future, whereas for people with elevated blood pressure, approximately **24 out of 100** will be affected.

Visual format



3.4. Variables and scales

3.4.1. Independent variable

Presentation format

Our independent variable was the presentation format that participants were exposed to. It consisted of statistical information being presented as percentages, natural frequencies or visually (icon arrays).

3.4.2. Demographic variables

Demographics

The demographic variables were limited to gender, age, and educational attainment. These were used for the descriptive statistics of our sample.

3.4.3. Control variable

Numeracy (SNS)

In order to measure participants' numeracy, we used the Subjective Numeracy Scale (SNS) developed by Fagerlin et al. (2007). It contains eight self-assessment items intended to measure quantitative ability. Whereas many studies use objective measures such as the Berlin Numeracy Test, the SNS can be more suitable for online surveys and is easier to administer and less of a burden for respondents (Fagerlin et al., 2007). The scale has strong correlation with objective numeracy and has been validated with ample predictive ability (Zikmund-Fisher, 2007). The scale includes items measuring self-assessed mathematical skill and respondents' preference for numerical information (Fagerlin et al., 2007). Items were measured on a 6-point Likert scale (some reverse coded) and a mean score was then calculated.

3.4.4. Dependent variables

Perceived likelihood (Absolute)

In order to measure the respondents' risk perception following the scenario descriptions, we adopted a scale from Janssen et al. (2012). In their article, the authors operationalized risk perception by distinguishing between cognitive and affective beliefs. Although they found that perceived affective likelihood might be a better predictor of future health behavior, they also stress that both constructs should not be considered in unison within the same study (Janssen et al., 2012). Since we were not examining future health behavior associated with risk perception, but rather participants' objective likelihood estimates, we choose to only include the cognitive dimension. Much like the original study, all questions were measured based on not performing the adaptive behavior (in our case, on not taking action to lower the blood pressure). This

was to ensure that all respondents based their estimations on the current condition of the patient (Janssen et al., 2012). Three items were measured on a 5-point Likert scale and a mean score was calculated and used for analysis. The questions were asking about the absolute risk (likelihood) of developing cardiovascular disease, explicitly stating that respondents were not to make comparison to the group with normal blood pressure.

Perceived likelihood (Relative)

Since the scenario contained information about the probabilities of developing cardiovascular disease for both normal and elevated blood pressure, we also set out to measure the perceived relative (comparative) likelihood. For this we adopted a scale used to measure the comparative likelihood in a conditional way (in the same way we measured absolute likelihood). Once again, only the cognitive dimension was considered and two items were measured on a 5-point Likert scale and a mean score was calculated. Items were developed based on earlier work from Janssen et al. (2011). For this variable we clearly stated that the questions were comparative in nature and asked for relative risk perceptions. The response alternatives referred to the risk of a person with elevated blood pressure to be *lower* or *higher* should they not take action to lower the blood pressure. See section 3.6.1. for discussion about the internal consistency of this variable and the revised version used in the final analysis.

Service performance (SERVPERF)

As a proxy for measuring consumer satisfaction, we selected a number of predictor items from the SERVPERF questionnaire that captures perceptions of service performance. Previous studies have found support for performance-based measures of service quality being superior to earlier constructs and that consumer satisfaction is a descendent of service quality (Cronin & Taylor, 1992). For the purpose of our study, we set up a revised SERVPERF scale with items that were relevant for the scenario at hand. The original scale measures five service dimensions with statements referring to: tangibles, reliability, responsiveness, assurance, and empathy (Jain & Gupta, 2004). The revised scale used in our study included six statements spread over the dimensions which were applicable based on the scenario with the exception of the tangibles category (as our scenario contained no descriptions of this). Items were measured on a 5-point Likert scale and a mean score was calculated and used for the final analysis.

Word-of-Mouth intention

An alternative measure we included to also capture consumer satisfaction was word-of-mouth intention. Not surprisingly, previous research suggest that a positive relationship exist between consumer satisfaction and word-of-mouth intention (Bolton & Drew, 1992). Our measure was inspired by the Likelihood-to-Recommend (LTR) question, where we simply asked the respondents to rate how likely it would be for them to recommend the physician in the scenario to a friend or colleague on a 7-point scale. Inspiration was taken from Boulding et al. (1993). This variable was mainly included for the purpose of testing the construct validity of our revised SERVPERF scale.

Understanding (RRI and Basic)

In an attempt to measure how well respondents were able to integrate probabilistic information based on the presentation format we took inspiration from Gigerenzer et al. (2008). In their article, the authors asked respondents to convert a percent to a proportion and then to translate a proportion to a percent (Gigerenzer et al., 2008). Our questions asked how many out of *1000 people*[1] with elevated blood pressure were estimated to develop cardiovascular disease in the future. This was a conscious choice in an attempt to gauge their basic understanding by having them convert from hundreds to thousands. Respondents were then to provide both percentage and frequency answers. For Understanding Basic a mean score for both questions was calculated and for sake of completeness we also constructed one variable for understanding percentage and one for understanding frequency. We also included one task testing respondents' ability to estimate the percentage increase in risk (Relative Risk Increase, RRI) between the group with normal and elevated blood pressure. Although the primary aim of these variables was to determine the effect of presentation format on understanding, the first two questions would be used to test the construct validity of the Subjective Numeracy Scale (SNS) and Relative Risk Increase (RRI) would be the primary variable testing their understanding. In a secondary analysis, we also investigated any potential impact of presentation format on the two basic understanding questions.

3.5. Data collection and analysis

3.5.1. Data collection

The survey was published and distributed on the 23rd of October 2023 and closed on the 31st of October. It was decided that a minimum of 50 answers was to be recorded for each treatment group which would add up to at least 150 valid responses. This number was chosen to ensure that our results had the potential to reach statistical significance. The questionnaire was distributed online through our own social channels. As was mentioned in section 3.1. this sampling strategy limits the generalizability of our findings and although we were able to capture a great deal of diversity in terms of respondents' demographics, it should not be viewed as a representative sample of the population (Bell et al., 2019). This convenience sampling method was however deemed most viable in light of the amount of time and resources available to us.

Only Swedish-speaking respondents were considered since providing the questionnaire in English could result in the scenario or questions being misunderstood. That being said, translating the established scales and items from English to Swedish came with its own risks but these were taken into careful consideration and multiple iterations were made to ensure congruency of the concepts.

[1] The risk statistics in the scenario were either presented in terms of percentages, or as x out of *100 people* for the frequency and visual group.

3.5.2. Data quality

In order to consider recorded responses for inclusion in the final analysis we set up a number of selection criteria. To start off, participants had to consent to the GDPR terms stated in the introduction of the survey. Any participant who did not consent was sent out of the survey immediately and no data was collected. Another requirement was that the whole survey had to be completed. By adding force response logic respondents were not able to move forward in the questionnaire if they had missed any items. Respondents also had to correctly answer the control question at the end but also pass the instructional manipulation check (IMC) included in the Service Performance block. Oppenheimer et al. (2009) claimed that the inclusion of an IMC has the potential to increase the statistical power and reliability of a dataset. This is achieved by making sure that participants are reading the instructions and involves embedding the IMC within a battery of questions. The respondents are then to provide a confirmation that they have read the instructions by answering in a specific way (Oppenheimer et al., 2009). In our case, this entailed answering "Completely disagree" to the statement "It is important to pay attention when completing this survey, please answer completely disagree".

3.5.3. Drop-out analysis

204 total responses had been recorded when the questionnaire was closed on the 31st of October. However, 7 respondents did not consent with the GDPR terms and were subsequently sent out of the survey. Out of the 197 remaining respondents, 35 did not complete the whole survey (progress was less than 100%) and were therefore excluded from the final analysis. Furthermore, 10 respondents failed the instructional manipulation check mentioned above. This left us with a sample size of 152 valid responses. All remaining respondents passed the last attention check.

3.5.4. Data analysis

The data was exported from Qualtrics to a SPSS v.29-compatible file. We made sure that all responses were exported with viewing order data for randomized surveys, which added a variable showing which experiment group each answer belonged to. Using SPSS, the data was cleaned and mean scores were calculated for multi-item variables. One-way analysis of variance (ANOVA) and Tukey's Honest Significant Difference (HSD) tests were performed after descriptive statistics had been generated.

3.6. Reliability and validity

3.6.1. Reliability

Reliability concerns the extent to which we can expect to get the same results from multiple measurements. If the applied measures yield consistent results, the reliability increases (Bell et al., 2019). For our multi-item variables, we used Cronbach's alpha to test their internal consistency. This was used to make sure that all items were coherent and related to the same constructs. For an acceptable level of internal reliability, we used the rule of thumb of an Alpha coefficient greater than 0.7 (Bell et al., 2019).

Table 1. Cronbach's Alpha coefficients

		Presentation Format			
Variable	No. of Items	Percentage	Frequency	Visual	Aggregated
Perceived Likelihood (Absolute)	3	.785	.831	.880	.835
Service Performance	6	.828	.849	.849	.842
Subjective Numeracy Scale	8	.855	.684	.822	.806

Cronbach's Alpha was calculated for each presentation format and on the aggregate level to compare internal consistency across the groups. Since respondents were exposed to different presentation formats in the beginning of the survey, we wanted to make sure that the treatment in itself did not influence responses. For all variables with more than three items the alpha coefficient was greater than 0.8 on the aggregate level, and almost greater than 0.7 for all the individual groups (with the exception of SNS for the frequency group which came close at 0.68).

Since two of our dependent variables were made up of two-item measures, we followed recommendations by Eisinga et al. (2013) and used the Spearman-Brown formula to estimate the reliability of these scales. The authors of that article emphasized that both Cronbach's Alpha and Pearson correlation are inappropriate to use for two-item scales and instead advocate the Spearman-brown formula (Eisinga et al., 2013).

Table 2. Spearman-Brown's coefficients

		Presentation Format			
Variable	No. of Items	Percentage	Frequency	Visual	Aggregated
Perceived likelihood (Relative)	2	-.02	.249	.315	.201
Understanding Basic	2	.884	.560	.727	.787

For the basic understanding variable, the aggregate coefficient indicated a good level of internal consistency. The frequency group stood out however, with a coefficient of only 0.56. Since this variable was only secondary in our analysis, we decided to include it but took this into consideration when doing the analysis. The low value could be due to the small sample size for each group or the low number of items included in the scale.

For the perceived likelihood (relative) variable the coefficient was unsatisfactory both on the aggregate and on the group level. This implies that the two items did not consistently measure the same underlying construct. This variable was made up of one question asking respondents to rate the relative risk of a person with elevated blood pressure (compared to normal blood pressure) and one statement which refered to the risk being "lower". For the sake of including this variable in our analysis, the statement item was removed and only the rating item was left.

3.6.2. Validity

To address questions about validity, that is to what extent our measures actually capture the intended concepts, we have done tests on construct validity. According to Bell et al. (2019) the validity of a measure can be tested by comparing it to other measures that are related to the same concepts. This is referred to as convergent validity. For the purpose of this thesis, the convergent validity of our SERVPERF variable was tested against the variable word-of-mouth intention. The subjective numeracy scale (SNS) was also tested against the basic understanding variable.

Table 3. Pearson correlation coefficients

		Presentation Format			
Variable 1	Variable 2	Percentage	Frequency	Visual	Aggregated
SERVPERF	W-O-M Intention	.576**	.832**	.676**	.713**
SNS	Basic Understanding	.26	.12	.27	.208**

* $p < .05$

** $p < .01$

Service Performance correlated well with Word-of-Mouth intention which indicates a good construct validity. A high rating on service performance should intuitively have a positive correlation with word-of-mouth intention since both concepts relate to consumer satisfaction (see section 3.4.4.). The subjective numeracy scale (SNS) had a positive and significant correlation with basic understanding on the aggregate level. For the individual groups this was not the case. Possible explanations for this could be that the group level analysis contains fewer data points compared to the aggregate level which affects the statistical power. On the other hand, there could be an aggregation bias where within-group variance is diluted on an aggregate level (which inflates the statistical power). There could also be an argument that the basic understanding variable (containing only two items) is not appropriate to use as a proxy for numeracy. Zikmund-Fisher et al. (2007) has a thorough review of the validity of the SNS-scale which shows a strong predictive ability on objective numeracy. For this reason, we consider the validity of the SNS-scale to be satisfactory but recognize that there might be unexplored, systematic differences in our sample compared to the original studies.

3.6.3. Survey evaluation

At the end of the questionnaire, respondents were asked to evaluate the survey and the scenario. Four items were measured on a 5-point Likert scale with statements referring to the clarity of the questions and the provided answer alternatives. One item concerned the realism of the scenario and the last asked if the questions tried to influence the respondents' answers.

Out of the 152 respondents, the majority found the questions and answer alternatives to be clear. Specifically, 87% agreed (55% strongly and 32% somewhat) that the questions were clearly formulated. Similarly, 93% agreed (59% strongly and 34% somewhat) that the answer options were clearly formulated.

As for the realism of the scenario, 89% of respondents agreed (62% strongly and 27% somewhat) that the scenario was realistic. Interestingly, a significant proportion of participants reported that the questions tried to influence their answers in some way (20% somewhat agreed and 10% strongly agreed). See Table 4 below for results.

Table 4. Survey evaluation

n = 152					
Questions	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
The questions were clearly formulated	0%	4%	9%	32%	55%
The answer options were clearly formulated	0%	2%	5%	34%	59%
The scenario was realistic	1%	1%	9%	27%	62%
The questions tried to influence your answers in some way.	30%	14%	26%	20%	10%

The participants who answered "Somewhat agree" and "Strongly agree" to this question were selected for a separate analysis to see if they had been exposed to any specific presentation format that might have biased their answers. The total number of individuals who belonged to this category was 46. It was however found that they were evenly spread out over all the presentation formats, with frequency being somewhat underrepresented (37% Percentage, 26% Frequency and 37% visual). This suggested that respondents' answers to this question were not systematically influenced by the presentation format.

4. Results and analysis

4.1. Descriptive statistics

4.1.1. Sample demographics

Out of the 152 valid responses a majority were males (61%) and the remaining were females (39%). The respondents had a mean age of 30. A large proportion of the sample had high school diplomas or equivalent (40%), 19 % had attended college or university for 1-3 years, and 34% for more than three years. Refer to Appendix 1 for a detailed breakdown of sample demographics.

Educational attainment, age, sex, and numeracy were similar for all three groups. A one-way ANOVA was conducted to control for this, and no significant differences were found between any of the groups. Refer to Appendix 2 for ANOVA output. Similarly, including numeracy as a covariate did not significantly impact the observed outcomes.

We recognize that our sample is not representative of the population which limits the generalizability of our findings. This was discussed under section 3.1. in the Methodology section.

4.2. Statistical analysis

The aim of this thesis was to examine how presentation format might influence important outcome variables related to risk perception, consumer satisfaction, and understanding. To achieve this, we used one-way ANOVA analysis, Tukey's HSD tests, and Pearson Correlation. A significance level of 5% ($p < .05$) is used throughout this section.

4.2.1. Analysis of variance (ANOVA) and Tukey's HSD tests

To determine if presentation format had any impact on the dependent variables, a one-way analysis of variance (ANOVA) was carried out. By comparing the means of our three groups for all dependent variables we were able to determine if at least one of them was significantly different from the others. See results in Table 5 below.

Table 5. One-way ANOVA results for dependent variables

Dependent Variables	Df	Mean Square	F	Significance	Effect Size (η^2)
Perceived Likelihood (Absolute)	2	.07	.11	.90	.01
Perceived Likelihood (Relative)	2	.764	3.407	.036*	.044
Service Performance	2	.04	.093	.91	.01
Word of Mouth Intention	2	.58	.26	.77	.004
UNDERSTANDING					
Understanding (RRI)	2	.839	3.440	.035*	.044
Understanding (Basic)	2	.28	2.47	.09	.032
Understanding (Percentage)	2	.32	2.18	.12	.028
Understanding (Frequency)	2	.28	2.09	.13	.027

* $p < .05$

For perceived likelihood (relative) and understanding (RRI), a significant difference in means was found indicating that presentation format had an effect on these two variables. To further analyze these differences, post-hoc tests (Tukey's HSD) were carried out. See Table 6 and 7 below for results.

Table 6. Tukey's HSD test results for Perceived Likelihood (Relative)

(I) Format	(J) Format	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Percentage	Frequency	-.20	.094	.08	-.43	.019
	Visual	-.02	.093	.97	-.20	.24
Frequency	Percentage	.20	.094	.08	-.02	.43
	Visual	.227*	.093	.050	.0001	.453
Visual	Percentage	.02	.093	.97	-.24	.20
	Frequency	-.227*	.096	.050	-.453	-.0001

* The mean difference is significant at the .05 level.

The results indicated that there were significant differences in means for frequency and visual format. More specifically, the mean difference tells us that, on average, individuals exposed to the frequency format perceived the relative likelihood of developing cardiovascular disease in the future as being higher than the visual group did. No significant differences were found for percentage format.

Table 7. Tukey's HSD test results for Understanding (RRI)

(I) Format	(J) Format	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Percentage	Frequency	-.255*	.098	.028	-.487	-.0228
	Visual	-.15	.097	.27	-.38	.08
Frequency	Percentage	.255*	.098	.028	.0228	.487
	Visual	.11	.01	.55	-.13	.34
Visual	Percentage	.15	.097	.27	-.08	.38
	Frequency	-.11	.01	.55	-.34	.13

* The mean difference is significant at the .05 level.

When participants were asked to estimate the percentage increase in risk (from having elevated blood pressure), we can see that there was a significant difference in understanding between frequency and percentage format. Only one alternative was correct. Individuals exposed to frequency format were better than percentage format at estimating RRI (based on the positive mean difference between these groups). No significant differences were found for visual format.

4.2.2. Significant difference in Perceived likelihood (Relative)

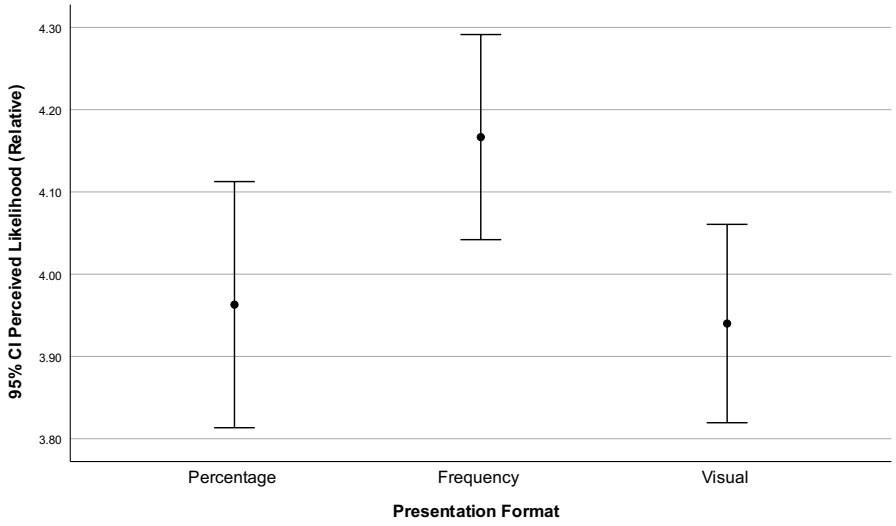
The revised version of Perceived Likelihood (Relative) consisted of one item measuring respondents' perception of relative risk on a 5-point Likert scale. The translated question was: "If the individual does not lower their blood pressure, the likelihood of developing cardiovascular disease in the future is:" 1) Much lower, to 5) Much higher, with 3) being neutral. The mean of the whole sample ($M = 4.02$, $SD = 0.48$) indicated that most respondents, on average, agreed that the risk of developing cardiovascular disease in the future was in fact higher if the individual did not lower their blood pressure.

Mean scores and standard deviations were:

- Percentage format group ($M = 3.96$, $SD = 0.55$).
- Frequency format group ($M = 4.17$, $SD = 0.44$).
- Visual format group ($M = 3.94$, $SD = 0.42$).

The one-way ANOVA revealed that there was a statistically significant difference in means between at least two groups, $F(2, 149) = 3.407$, $p = .036$. The effect size ($\eta^2 = .044$) indicated a small to medium effect (Cohen, 1988). This implies that 4.4% of the variance was explained by the presentation format. Tukey's HSD Test for multiple comparisons found that the difference was significant between frequency format and visual format ($p = .05$). On average, frequency format resulted in participants perceiving the relative likelihood of developing cardiovascular disease in the future as being higher than the visual format did. No significant differences were found for percentage format. See figure 1 below for error bars.

Figure 1. Error Bars with 95% Confidence Interval for mean of Perceived Likelihood (Relative)



Note: 1 Observe that scales are not set in reference to variable measurement range. The range has been chosen to clarify mean differences which otherwise may have been difficult to see. Refer to Appendix 4 for non-adjusted figure.

4.2.3. Significant difference in Understanding (RRI)

The variable understanding (RRI) asked respondents to estimate the percentage increase in risk between an individual with elevated blood pressure and an individual with normal blood pressure. Four alternatives were provided with only one being correct (the RRI was 70% in the scenario, an increase from 14% to 24%). Before providing an answer, respondents were reminded of the risk statistics they had been presented to in the scenario (with the presentation format corresponding to which group they had been randomly assigned to from the beginning).

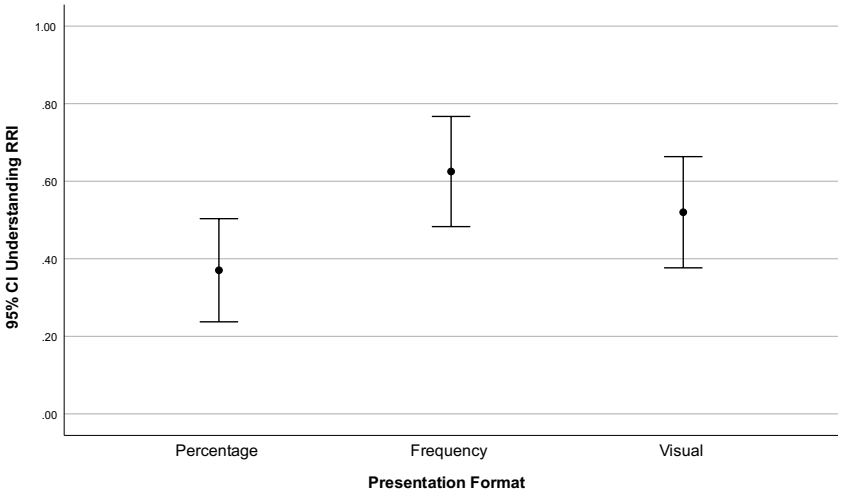
Mean scores and standard deviations were:

- Percentage format group ($M = 0.37$, $SD = 0.49$).
- Frequency format group ($M = 0.63$, $SD = 0.49$).
- Visual format group ($M = 0.52$, $SD = 0.50$).

The whole sample had a mean of 0.50 ($SD = 0.50$) indicating that, on average, 50% gave a correct answer. The one-way ANOVA revealed that there was a statistically significant difference between at least two groups, $F(2, 149) = 3.440$, $p = .035$. The estimated effect size ($\eta^2 = .044$) indicated a small to medium effect (Cohen, 1988). Tukey's HSD post hoc test showed that the frequency format group scored significantly higher than the percentage format group ($p = .028$). On average, 63% got the correct answer when being exposed to frequency format, and only 37% were correct when exposed to percentage format. It is worth pointing out that a sizeable portion of

individuals exposed to percentage format (52%) answered that the RRI was 10%. In the frequency format group this proportion was 21% and for visual 38%. Although this is the percentage point increase, it was not asked for in the question. Refer to Appendix 6 for complete overview of answer distribution. See figure 2 below for error bars.

Figure 2. Error Bars with 95% Confidence Interval for mean of Understanding (RRI)



4.2.4. No significant difference in Perceived Likelihood (Absolute)

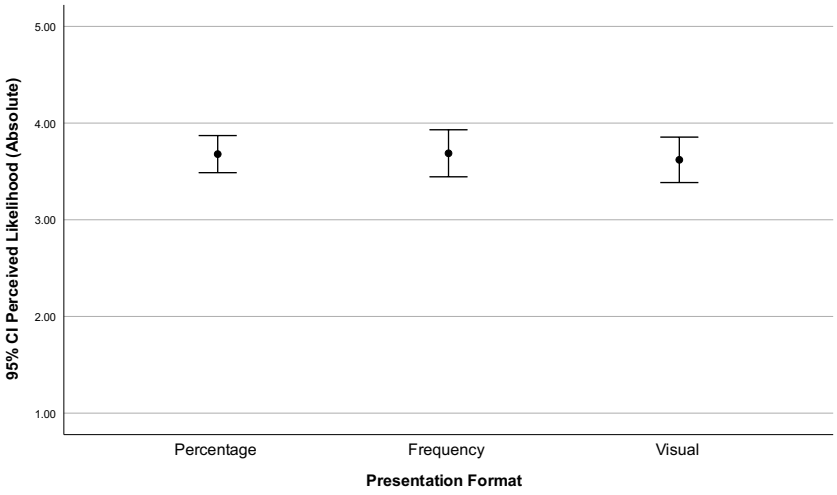
Respondents' perception of the absolute likelihood of developing cardiovascular disease in the future for an individual with elevated blood pressure (that is, irrespective of the baseline risk for the normal group) was based on two items adopted from Janssen et al. (2012). Items were measured on 5-point Likert scale and a mean score was calculated. The whole sample had a mean of 3.66 (SD = 0.78) indicating that most respondents perceived the absolute risk as being modest. No significant differences in means were found for any of the groups, $F(2, 149) = 0.11$, $p = .90$.

Mean scores and standard deviations were:

- Percentage format group (M = 3.68, SD = 0.70).
- Frequency format group (M = 3.69, SD = 0.84).
- Visual format group (M = 3.52, SD = 0.83).

See figure 3 below for error bars.

Figure 3. Error bars with 95% Confidence Interval for mean of Perceived Likelihood (Absolute)



4.2.5. No significant difference in Service Performance or W-O-M Intention

As two alternative measures relating to consumer satisfaction, service performance and word-of-mouth intention were included in the analysis (the former assumed to be an antecedent to consumer satisfaction and the latter having a positive relationship with it). The one-way ANOVA revealed no significant differences in service performance, $F(2, 149) = 0.093$, $p = .91$. The mean for the sample as a whole was 4.15 ($SD = 0.63$) on a 5-point scale indicating a positive perception of service performance.

Mean scores and standard deviations were:

- Percentage format group ($M = 4.15$, $SD = 0.53$).
- Frequency format group ($M = 4.12$, $SD = 0.70$)
- Visual format group ($M = 4.18$, $SD = 0.67$).

Similarly, no significant differences were found in w-o-m intention, $F(2, 149) = 0.26$, $p = .77$. The variable was measured on a 7-point scale with a sample mean of 5.01 ($SD = 1.04$) indicating that most respondents were likely to recommend the physician in the scenario.

Mean scores and standard deviations were:

- Percentage format group ($M = 5.04$, $SD = 0.78$).
- Frequency format group ($M = 5.06$, $SD = 1.28$).
- Visual format group ($M = 4.92$, $SD = 1.05$).

See figure 4 and 5 below for error bars.

Figure 4. Error Bars with 95% Confidence Interval for mean of Service Performance

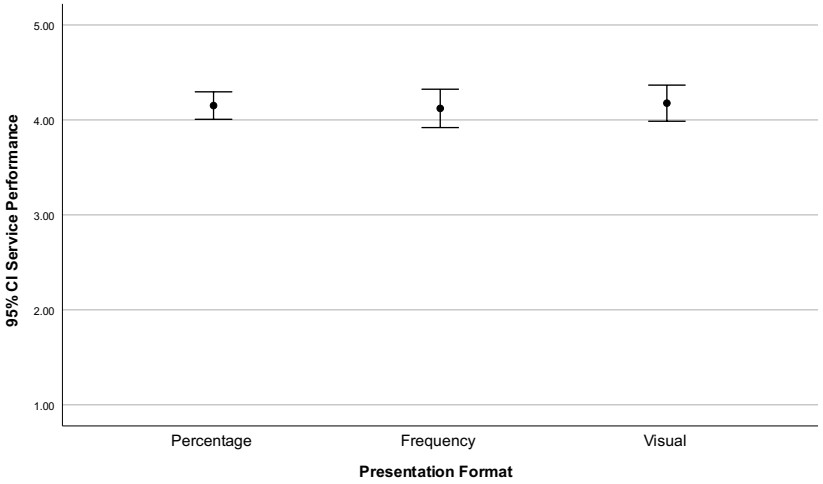
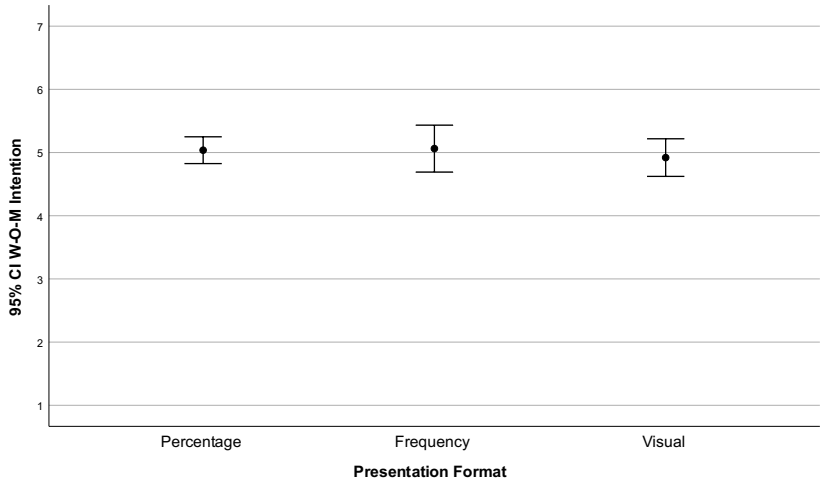


Figure 5. Error Bars with 95% Confidence Interval for mean of W-O-M Intention



4.2.6. No significant difference in Basic Understanding

The Basic Understanding variable was inspired by Gigerenzer et al. (2008) and included one item measuring respondents' ability to estimate the number of individuals with elevated blood pressure expected to develop cardiovascular disease in the future, and one item asking them to estimate the percentage of individuals. Results revealed no significant differences in accuracy between groups, $F(2, 149) = 2.47$, $p = .09$. The same was true for the individual components of the measure (that is looking at Understanding Frequency and Understanding Percentage in isolation).

Mean scores and standard deviations were:

- Percentage format group ($M = 0.84$, $SD = 0.35$),
- Frequency format group ($M = 0.75$, $SD = 0.39$)
- Visual format group ($M = 0.90$, $SD = 0.27$).

The mean for the whole sample was 0.83 ($SD = 0.34$) indicating that, on average, respondents answered almost both questions correctly (a score of 0.50 would indicate one correct answer out of two). Visual format yielded the highest proportion of correct answers for all basic understanding measures although no statistical significance was found. See table 8 below for overview.

Table 8. Proportion of correct answers

	Presentation Format		
Variable	Percentage	Frequency	Visual
Basic	84%	75%	90%
Percentage	85%	73%	88%
Frequency	83%	77%	92%

4.3. Variable correlations

There were a number of noteworthy correlations based on their significance level and their coefficients. For a complete overview refer to Appendix 5. For the aggregate group, understanding (RRI) and perceived likelihood (Absolute) had a significant negative correlation ($r = -.213$, $p < .01$). This indicates that perceiving the absolute likelihood of developing cardiovascular disease as higher was associated with lower understanding of RRI. As discussed in the methodology section, word-of-mouth intention correlated strongly with SERVPERF ($r = .713$, $p < .01$) which supports the construct validity of our SERVPERF variable. Numeracy (SNS) correlated positively and significantly with all understanding measures. There was also a significant positive correlation between understanding (RRI) and understanding (Basic), as one would expect ($r = .223$, $p < .01$).

4.4. Summary of analysis

A one-way ANOVA revealed that there was a significant effect of presentation format on perceived likelihood (relative), $F(2, 149) = 3.407$, $p = .036$. The effect size ($\eta^2 = .044$) indicated a small to medium effect. Tukey's HSD Test for multiple comparisons revealed that individuals exposed to frequency format estimated the relative likelihood of developing cardiovascular disease in the future as being higher than the visual group did ($p = .05$).

A significant effect was also found on understanding (RRI), $F(2, 149) = 3.440$, $p = .035$. The effect size ($\eta^2 = .044$) again, indicated a small to medium effect. Tukey's HSD found that the effect was significant between frequency format and percentage format ($p = .028$). The mean difference between the two groups revealed that individuals exposed to frequency format had higher accuracy when estimating the percentage increase in risk. On average, 63% got the correct answer when being exposed to frequency format, and only 37% were correct when exposed to percentage format.

For all other variables and comparison sets, no significant effects were found. See Appendix 3 for a complete summary statistics table.

4.5. Hypotheses summary

For the basic guiding hypotheses concerning understanding, empirical support was found for **H1** but not for **H2.** Although no significant effects were found for basic understanding, this variable was mainly included to test the construct validity of the SNS-scale. Refer to section 3.4.4. for rationale. The primary variable used for the hypotheses testing was understanding (RRI).

Table 9 Summary of hypotheses

H1	Individuals exposed to frequency format will have better understanding compared to individuals exposed to percentage format.	Supported
H2	Individuals exposed to visual format will have better understanding compared to individuals exposed to percentage format.	Not supported

5. Discussion

5.1. Conclusions and implications

5.1.1. Perceived likelihood (Relative)

The results revealed that presentation format had a significant effect on the perceived relative likelihood of developing cardiovascular disease in the future. This was however only the case when comparing frequency format with visual format. It was found that the former resulted in higher relative risk perceptions compared to the latter. A possible explanation that was discussed in Section 2.3. is that humans are more adept at understanding and processing natural frequencies. To the extent that we can say that a higher relative risk perception is "more correct", it might simply be the case that individuals exposed to natural frequencies have an easier time to process the information. What is interesting is that the visual format with icon arrays also included natural frequencies, although the number of individuals not at risk is also spelled out directly and shown on the graphical illustration. It has been suggested in previous research that using icon arrays could in fact make the number of unaffected people stand out (Galesic et al., 2009). For this very reason, it could be that individuals in the visual group reported the relative risk as being lower compared to the frequency format group.

Although modest, the effect size ($\eta^2 = .044$) indicated a small to medium effect. Cohen (1988) noted that meaningful effects within behavioral science might in fact be in this small order of magnitude. He further emphasized that achieving very high values is not common outside of natural sciences (Cohen, 1988). The implication from this finding is not clear-cut. It is not obvious that perceiving the relative risk as being higher is more correct or even desirable. Assuming that risk communication should be transparent and not try to deceive consumers, the application should be adapted based on the situation. In this specific scenario, a higher perception of relative risk seems to be related to better understanding of relative risk increase (se section 5.1.2. below).

An important note is that although there was a statistically significant difference between frequency format and visual format, the same was not true for percentage format. This could mean that percentage format is similar to both frequency and visual format, or at least not different enough to reach statistical significance. It should be pointed out that when comparing frequency format with percentage format, the significance level almost reached the threshold ($p = 0.08$) with a positive mean difference. If a larger sample size would have been used, it is possible that statistical significance would have been reached. Again, the mean difference indicated that the frequency format group had higher perceptions of relative likelihood than percentage format.

5.1.2. Understanding (RRI)

Accuracy of understanding relative risk increase was significantly influenced by presentation format. More specifically, individuals exposed to frequency format were better at estimating the percentage increase compared to individuals exposed to percentage format. This was in line with previous research which has suggested that natural frequencies improves people's statistical reasoning (McDowell & Jacobs, 2017; Gigerenzer & Hoffrage, 1995). Interestingly, frequency format elicited higher perceptions of relative risk *and* higher accuracy of understanding relative risk increase. However, an important note that was made in the results section is that a considerable proportion of individuals in the percentage format group (52%) and visual format group (38%) seemed to have difficulty differentiating between percentages and percentage points. These individuals estimated that the percentage increase was 10% (the correct answer was 70%). The proportion of individuals in the frequency group that had this same difficulty was only 21%. The percentage point increase was indeed 10 (the difference between 24% and 14%), but the *percentage increase* was explicitly asked for. This finding is not trivial, as it might suggest that rather than frequency format increasing relative risk understanding, it could be that natural frequencies facilitates differentiation between percentages and percentage points. Since numeracy was similar at baseline for all groups, this effect should be attributable to the presentation format. Although this finding is important in itself, one should be conservative before jumping to any conclusion that frequency format increases understanding of relative risk increase based on this study alone. Nevertheless, the implication remains that frequency format seems to facilitate statistical reasoning in a broad sense.

Again, no significant differences were found for visual format (compared to both frequency and percentage format). We cannot confidently say that the visual format group was better or worse at understanding RRI compared to the frequency or percentage format groups. In a similar vein, it is not certain that frequency format was better than visual format in increasing understanding. This was somewhat expected, as previous studies have found conflicting evidence of the efficacy of icon arrays (Galesic et al., 2009; Ruiz et al., 2013).

5.1.3. Service Performance and Word-of-Mouth Intention

Presentation format did not significantly impact service performance or word-of-mouth intention. Although empirical support of no effects are valuable in and of themselves, a discussion about the appropriateness of these measures might be warranted too. We did receive some comments about it being difficult to rate service performance based solely on the written scenario. In hindsight, we do acknowledge that a written scenario in a questionnaire might lack the necessary depth of a real interaction that is needed to rate service performance or estimate intentions to engage in word of mouth. For this reason, we are careful in suggesting any implications. Granted that presentation format would

have had a larger impact on understanding, it might be hypothesized that consumer satisfaction would increase as well (and therefore service performance and W-O-M Intention too). For this reason, it would be inappropriate to suggest that presentation format is of no importance for consumer satisfaction. In a scenario requiring more complex understanding, presentation format might be more significant.

5.1.4. Perceived Likelihood (Absolute) and Understanding (Basic)

Interestingly, the group exposed to visual format had the highest proportion of correct answers for all basic understanding measures although no statistical significance was found. One might have speculated a priori that the percentage format group would have the highest accuracy on the item relating to percentages, and that frequency format would score best on the item relating to frequencies. For both items the visual group was however superior (see Table 8 in the results section). Again, no statistical significance was found and this could simply be due to random chance. Likewise, presentation format had no significant impact on perceived likelihood (absolute). It could be that statistics stated in absolute terms are easier to interpret and, as a result, perceptions will be unaffected by presentation format. The advantage of natural frequencies in statistical reasoning implies nothing about how subjects should perceive absolute risk. Nevertheless, this variable was interesting to include since there has been evidence of, for example, icon arrays increasing the perceived seriousness of risk (Galesic et al., 2009). For this reason, it could be speculated that this perception bias would spill over into perceiving the absolute risk as being higher as well. No evidence was however found for this. The proposed implication is that presentation format is of less importance when communicating absolute risk. In a similar vein, basic understanding of risk communication also seems unaffected by presentation format. This suggests that context is crucial when evaluating an appropriate presentation format. For risk communication involving more cognitive processing and statistical reasoning, frequency format seems most appropriate. For simpler communication about risk, the format is of less importance.

5.2. Limitations

We acknowledge that this thesis has limitations, some of which have been touched upon in previous sections, and some that have yet to be discussed. First and foremost, the data collection strategy was a form of convenience sampling (limited to our own social networks) and although we did achieve some diversity in terms of respondents' demographics, it should not be viewed as a true representation of the population. This limits the generalizability of our findings.

Secondly, our measures of both perceived likelihood (relative) and understanding (RRI) consisted of one only item each. Perceived likelihood (relative) was reduced from two

items to one since we found issues with the internal reliability of the original scale (refer to section 3.6.1. for more details). It is widely recognized that more items lead to better construct representation and using too few is considered problematic (Eisinga et al., 2013; Marsh et al., 1998). Since these were the only variables where significant differences were found, even more prudence is called for when interpreting the results.

Thirdly, an important phenomenon related to risk communication is the concept of risk targets and the interplay of comparative optimism. When subjects have been asked to rate the risk for different activities there have been considerable inconsistencies depending on the specified risk target. Whether they are instructed to rate the risk to themselves, their family, or people in general there is a strong tendency to rate the risk to others as being significantly higher than to themselves (Sjöberg, 2000). This phenomenon, coined comparative optimism, describes the tendency of individuals to believe that negative events are more likely to happen to others than to themselves (Chambers et al., 2003). This has important implications for our study since most variables had a third person (the individual they read the scenario about) as specified risk target. This experiment design feature was chosen out of consideration for GDPR (in order to not collect sensitive data related to individual health status) and ethical concerns. Consequently, it is possible that respondents rated both the absolute and the relative risk as being higher than they would have if the scenario referred directly to themselves. The same is true for the other variables which referred to the individual in the scenario and not to the respondent. For this reason, the results could have been different if the risk target was specified as the respondents themselves and the scenario was written in a second-person narrative (the "you" perspective).

Lastly, it must be mentioned that communicating risks about other illnesses than cardiovascular disease might drastically change the results. Although severe, perceptions of cardiovascular disease might be significantly different than perceptions of, for example, cancer. It could be hypothesized that the choice of disease will evoke different emotional reactions which could very well impact the results. The same might be said about diseases considered "less severe". For this reason, the key findings of this thesis might not be applicable to other contexts.

5.2.1. Improvements for future research

If we were to redo this study, all the points above would be taken into consideration. We would have used true probability sampling to strengthen the generalizability of our findings. Furthermore, we would make sure that all variables included in the analysis contained a sufficient number of items. It would also be interesting to think about changing the risk target and designing an experiment that referred directly to the respondents. To gain deeper insights about understanding, it would be valuable to also ask for a comment on the respondents' thought processes when estimating RRI. It is possible that the significant effect on this variable was due to confusion of percentage

increase and percentage points. To capture consumer satisfaction, we would look at other ways of measuring this even when the scenario might lack the nuances of a real encounter. We would also go beyond just communicating health risk and see if different effects are found in other domains (for example for financial or insurance products).

5.3. Conclusions and final words

The purpose of this thesis was to examine how presentation format might impact important outcome variables related to risk communication. It is our hope that we were able to shed some light on the psychological processes that underpin risk perception, understanding and consumer satisfaction. In the best of worlds, this might help inform public and private entities on the most appropriate "language of risk". Results reveal that all three presentation formats are suitable when communicating basic and absolute risk, but frequency format seems most appropriate when communicating risks that require more cognitive processing and statistical reasoning. With that being said, more research is needed to reach a definitive answer of which presentation format is most appropriate for other situations.

6. References

Apter, A. J., Paasche-Orlow, M. K., Remillard, J. T., Bennett, I. M., Ben-Joseph, E. P., Batista, R. M., Hyde, J., & Rudd, R. E. (2008). Numeracy and Communication with Patients: They Are Counting on Us. *Journal of General Internal Medicine*, *23*(12), 2117–2124. https://doi.org/10.1007/s11606-008-0803-x

Bell, E., Bryman, A., & Harley, B. (2019). Business Research Methods (5th ed.). Oxford University Press.

Bolton, R. N., & Drew, J. H. (1992). Mitigating the effect of service encounters. *Marketing Letters*, *3*(1), 57–70. https://doi.org/10.1007/BF00994081

Boulding, W., Kalra, A., Staelin, R., & Zeithaml, V. A. (1993). A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions. *Journal of Marketing Research*, *30*(1), 7. https://doi.org/10.2307/3172510

Chambers, J. R., Windschitl, P. D., & Suls, J. (2003). Egocentrism, Event Frequency, and Comparative Optimism: When what Happens Frequently is "More Likely to Happen to Me." *Personality and Social Psychology Bulletin*, *29*(11), 1343–1356. https://doi.org/10.1177/0146167203256870

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587

Cronin, J. J., & Taylor, S. A. (1992). Measuring Service Quality: A Reexamination and Extension. *Journal of Marketing*, *56*(3), 55. https://doi.org/10.2307/1252296

DiClemente, R. J., & Jackson, J. M. (2017). Risk Communication. In *International Encyclopedia of Public Health* (pp. 378–382). Elsevier. https://doi.org/10.1016/B978-0-12-803678-5.00389-1

Eisinga, R., Grotenhuis, M. te, & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

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. *Medical Decision Making*, *27*(5), 672–680. https://doi.org/10.1177/0272989X07304449

Fischhoff, B. (2011). *Communicating Risks and Benefits: An Evidence Based User's Guide*. U.S. Deptartment of Health and Human Services.

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology*, *28*(2), 210–216. https://doi.org/10.1037/a0014474

Garcia-Retamero, R., & Dhami, M. K. (2011). Pictures speak louder than numbers: on communicating medical risks to immigrants with limited non-native language proficiency. *Health Expectations*, *14*(s1), 46–57. https://doi.org/10.1111/j.1369-7625.2011.00670.x

Garcia-Retamero, R., & Galesic, M. (2010). How to Reduce the Effect of Framing on Messages About Health. *Journal of General Internal Medicine*, *25*(12), 1323–1329. https://doi.org/10.1007/s11606-010-1484-9

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest*, *8*(2), 53–96. https://doi.org/10.1111/j.1539-6053.2008.00033.x

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. https://doi.org/10.1037/0033-295X.102.4.684

Hedberg, B., Wijk, H., Andersson Gäre, B., & Petersson, C. (2022). Shared decision-making and person-centred care in Sweden: Exploring coproduction of health and social care services. *Zeitschrift Für Evidenz, Fortbildung Und Qualität Im Gesundheitswesen*, *171*, 129–134. https://doi.org/10.1016/j.zefq.2022.04.016

Hertwig, R., Andrea Zangerl, M., Biedert, E., & Margraf, J. (2008). The public's probabilistic numeracy: How tasks, education and exposure to games of chance shape it. *Journal of Behavioral Decision Making*, *21*(4), 457–470. https://doi.org/10.1002/bdm.611

Hoerger, M. (2010). Participant Dropout as a Function of Survey Length in Internet-Mediated University Studies: Implications for Study Design and Voluntary Participation in Psychological Research. *Cyberpsychology, Behavior, and Social Networking*, *13*(6), 697–700. https://doi.org/10.1089/cyber.2009.0445

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating Statistical Information. *Science*, *290*(5500), 2261–2262. https://doi.org/10.1126/science.290.5500.2261

Jain, S. K., & Gupta, G. (2004). Measuring Service Quality: Servqual vs. Servperf Scales. *Vikalpa: The Journal for Decision Makers*, *29*(2), 25–38. https://doi.org/10.1177/0256090920040203

Janssen, E., van Osch, L., de Vries, H., & Lechner, L. (2011). Measuring risk perceptions of skin cancer: Reliability and validity of different operationalizations. *British Journal of Health Psychology*, *16*(1), 92–112. https://doi.org/10.1348/135910710X514120

Janssen, E., van Osch, L., Lechner, L., Candel, M., & de Vries, H. (2012). Thinking versus feeling: Differentiating between cognitive and affective components of perceived cancer risk. *Psychology & Health*, *27*(7), 767–783. https://doi.org/10.1080/08870446.2011.580846

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), 341–350. https://doi.org/10.1037/0003-066X.39.4.341

Kamrin, M. A. (2014). Risk Communication. In *Encyclopedia of Toxicology* (pp. 174–176). Elsevier. https://doi.org/10.1016/B978-0-12-386454-3.00430-9

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is More Ever Too Much? The Number of Indicators per Factor in Confirmatory Factor Analysis. *Multivariate Behavioral Research*, *33*(2), 181–220. https://doi.org/10.1207/s15327906mbr3302_1

McDowell, M., & Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin*, *143*(12), 1273–1312. https://doi.org/10.1037/bul0000126

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Recchia, G., Lawrence, A. C. E., & Freeman, A. L. J. (2022). Investigating the presentation of uncertainty in an icon array: A randomized trial. *PEC Innovation*, *1*, 100003. https://doi.org/10.1016/j.pecinn.2021.100003

Rietz, H., Pennlert, J., Nordström, P., & Brunström, M. (2023). Blood Pressure Level in Late Adolescence and Risk for Cardiovascular Events : A Cohort Study. *Annals of Internal Medicine*. https://doi.org/10.7326/M23-0112

Rothman, R. L., Montori, V. M., Cherrington, A., & Pignone, M. P. (2008). Perspective: The Role of Numeracy in Health Care. *Journal of Health Communication*, *13*(6), 583–595. https://doi.org/10.1080/10810730802281791

Ruiz, J. G., Andrade, A. D., Garcia-Retamero, R., Anam, R., Rodriguez, R., & Sharit, J. (2013). Communicating global cardiovascular risk: Are icon arrays better than numerical estimate *Patient Education and Counseling*, *93*(3), 394–402. https://doi.org/10.1016/j.pec.2013.06.026

Sjöberg, L. (2000). Factors in Risk Perception. *Risk Analysis*, *20*(1), 1–12. https://doi.org/10.1111/0272-4332.00001

Stovitz, S. D., & Shrier, I. (2013). Medical decision making and the importance of baseline risk. *British Journal of General Practice*, *63*(616), e795–e797. https://doi.org/10.3399/bjgp13X674585

Zikmund-Fisher, B. J., Smith, D. M., Ubel, P. A., & Fagerlin, A. (2007). Validation of the Subjective Numeracy Scale: Effects of Low Numeracy on Comprehension of Risk Communications and Utility Elicitations. *Medical Decision Making*, *27*(5), 663–671. https://doi.org/10.1177/0272989X07303824

7. Appendices

Appendix 1. Sample demographics

	N		Percentage
Variables	152	n	100%
GENDER			
Male		93	61.2
Female		59	38.8
AGE (*M* = 30)			
18 – 24		70	46.1
25 – 29		40	26.3
30 – 39		13	8.6
40 – 49		7	4.6
> 50		22	14.5
EDUCATIONAL ATTAINMENT			
Elementary school or equivalent		4	2.6
High school diploma		61	40.1
Vocational education		6	3.9
College or university (1–3 years)		29	19.1
College or university (>3 years)		52	34.2
PRESENTATION FORMAT			
Percentage format		54	35.5
Frequency format		48	31.6
Visual format		50	32.9

Appendix 2. One-way analysis of variance (ANOVA) for sample demographics and numeracy

Variable	Mean Square	F	Df	Significance	Effect Size (η^2)
DEMOGRAPHICS					
Age	90.70	.50	2	.61	.01
Gender	.17	.73	2	.49	.01
Educational Attainment	1.46	.77	2	.47	.01
CONTROL					
Subjective Numeracy Scale	.14	.16	2	.84	.02

Appendix 3. Summary statistics table

Panel A: Descriptive Statistics for All Presentation Formats (n = 152)

Variable	Mean	SD	Median	Minimum	Maximum
Perceived Likelihood (Absolute)	3.66	.78	4.00	1.67	5.00
Perceived Likelihood (Relative)	4.02	.48	4.00	2.00	5.00
Service Performance	4.15	.63	4.17	1.83	5.00
Word of Mouth Intention	5.01	1.04	5.00	1.00	7.00
Understanding RRI	.50	.50	.50	.00	1.00
Understanding Basic	.83	.34	1.00	.00	1.00
Understanding Percentage	.82	.38	1.00	.00	1.00
Understanding Frequency	.84	.37	1.00	.00	1.00
Subjective Numeracy Scale	4.34	.91	4.44	1.88	6.00

Panel B: Descriptive Statistics for Percentage Format (n = 54)

Variable	Mean	SD	Median	Minimum	Maximum
Perceived Likelihood (Absolute)	3.68	.70	4.00	1.67	4.33
Perceived Likelihood (Relative)	3.96	.55	4.00	2.00	5.00
Service Performance	4.15	.53	4.00	3.00	5.00
Word of Mouth Intention	5.04	.78	5.00	3.00	7.00
Understanding RRI	.37	.49	.00	.00	1.00
Understanding Basic	.84	.35	1.00	.00	1.00
Understanding Percentage	.85	.36	1.00	.00	1.00
Understanding Frequency	.83	.38	1.00	.00	1.00
Subjective Numeracy Scale	4.29	1.02	4.25	1.88	6.00

Panel C: Descriptive Statistic for Frequency Format (n = 48)

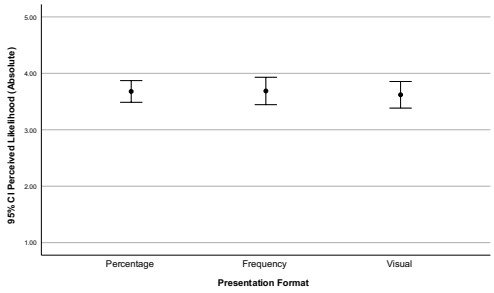
Variable	Mean	SD	Median	Minimum	Maximum
Perceived Likelihood (Absolute)	3.69	.84	4.00	1.67	5.00
Perceived Likelihood (Relative)	4.17	.43	4.00	3.00	5.00
Service Performance	4.12	.70	4.33	1.83	5.00
Word of Mouth Intention	5.06	1.28	5.00	1.00	7.00
Understanding RRI	.63	.49	1.00	.00	1.00
Understanding Basic	.75	.39	1.00	.00	1.00
Understanding Percentage	.73	.45	1.00	.00	1.00
Understanding Frequency	.77	.42	1.00	.00	1.00
Subjective Numeracy Scale	4.39	.75	4.38	2.75	5.63

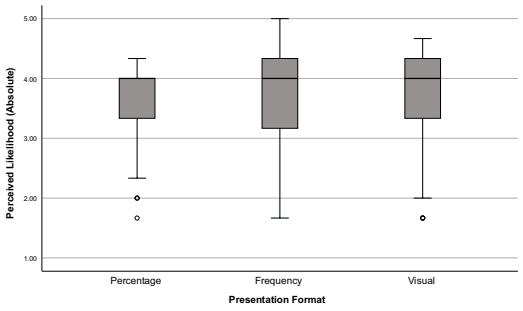
Panel D: Descriptive Statistic for Visual Format (n = 50)

Variable	Mean	SD	Median	Minimum	Maximum
Perceived Likelihood (Absolute)	3.62	.83	4.00	1.67	4.67
Perceived Likelihood (Relative)	3.94	.42	4.00	2.00	5.00
Service Performance	4.18	.67	4.25	2.50	5.00
Word of Mouth Intention	4.92	1.05	5.00	2.00	7.00
Understanding RRI	.52	.50	1.00	.00	1.00
Understanding Basic	.90	.27	1.00	.00	1.00
Understanding Percentage	.88	.33	1.00	.00	1.00
Understanding Frequency	.92	.27	1.00	.00	1.00
Subjective Numeracy Scale	4.36	.94	4.63	2.25	6.00

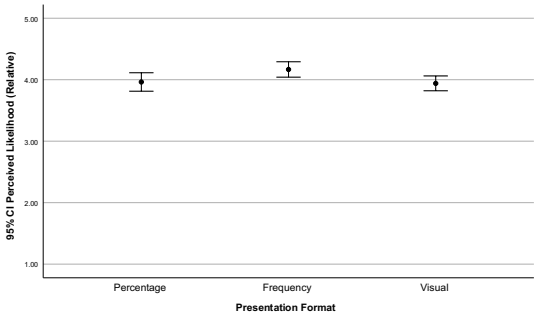
Appendix 4. Error Bars with 95% Confidence Interval for mean and Boxplots

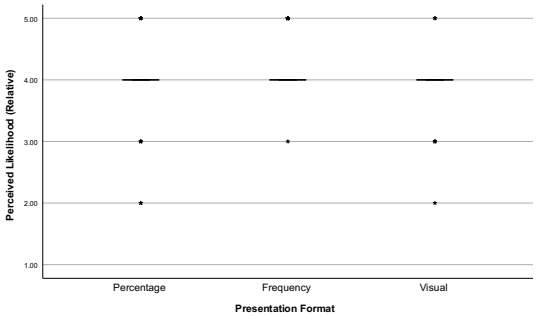
Perceived Likelihood (absolute)





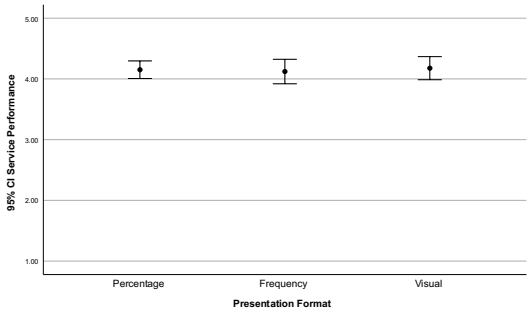
Perceive Likelihood (relative)

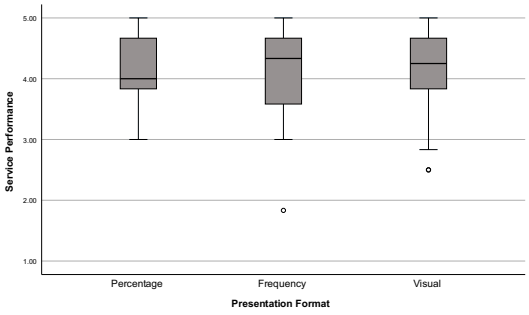




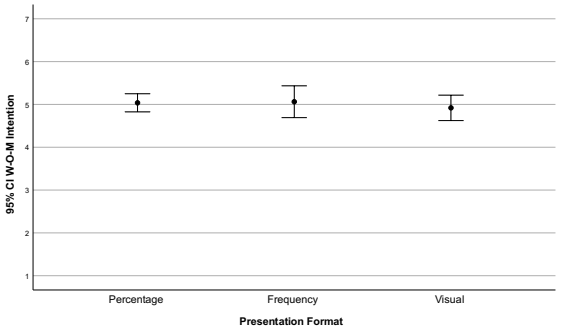
Note: 1 Unadjusted range for error bars

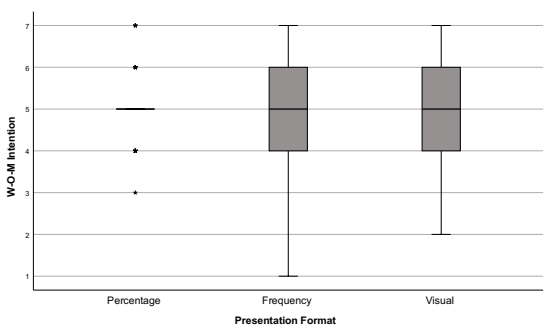
Service Performance





Word of Mouth Intention





Appendix 5. Pearson correlation matrix

PANEL A Correlations between variable for all presentation formats

	M	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Perceived Likelihood (Absolute)	3.66	0.78	1.00	-.01	.14	.10	-.213**	-.177*	-.171*	-.15	-.16
2. Perceived Likelihood (Relative)	4.02	0.48	-.01	1.00	.02	.03	.12	.04	.06	.02	.11
3. Service Performance	4.15	0.63	.14	.02	1.00	.713**	-.04	-.12	-.15	-.05	.09
4. W-O-M Intention	5.01	1.04	.10	.03	.713**	1.00	-.12	-.166*	-.197*	-.10	.03
5. Understanding RRI	0.50	0.50	-.213**	.12	-.04	-.12	1.00	.223**	.224**	.180*	.304**
6. Understanding Basic	0.83	0.34	-.177*	.04	-.12	-.166*	.223**	1.00	.912**	.903**	.208**
7. Understanding Percentage	0.82	0.38	-.171*	.06	-.15	-.197*	.224**	.912**	1.00	.648**	.207*
8.Understanding Frequency	0.84	0.37	-.15	.02	-.05	-.10	.180*	.903**	.648**	1.00	.171*
9. Subjective Numeracy Scale	4.34	0.91	-.16	.11	.09	.03	.304**	.208**	.207*	.171*	1.00

PANEL B Correlations between variables for Percentage format

	M	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Perceived Likelihood (Absolute)	3.68	0.70	1.00	-0.05	0.05	-0.17	-.270*	-0.15	-0.14	-0.14	-.322*
2. Perceived Likelihood (Relative)	3.96	0.55	-0.05	1.00	0.03	0.00	-0.02	-0.03	-0.03	-0.03	0.03
3. Service Performance	4.15	0.53	0.05	0.03	1.00	.576**	-0.11	-0.03	-0.05	-0.01	0.12
4. W-O-M Intention	5.04	0.78	-0.17	0.00	.576**	1.00	-0.14	0.02	0.02	0.02	0.23
5. Understanding RRI	0.37	0.49	-.270*	-0.02	-0.11	-0.14	1.00	0.13	0.10	0.14	.314*
6. Understanding Basic	0.84	0.35	-0.15	-0.03	-0.03	0.02	0.13	1.00	.944**	.949**	0.26
7. Understanding Percentage	0.85	0.36	-0.14	-0.03	-0.05	0.02	0.10	.944**	1.00	.793**	0.26
8.Understanding Frequency	0.83	0.38	-0.14	-0.03	-0.01	0.02	0.14	.949**	.793**	1.00	0.24
9. Subjective Numeracy Scale	4.29	1.02	-.322*	0.03	0.12	0.23	.314*	0.26	0.26	0.24	1.00

PANEL C Correlations between variable for Frequency format

	M	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Perceived Likelihood (Absolute)	3.69	0.84	1.00	-.09	.22	.14	-.448**	-.400**	-.381**	-.325*	-.04
2. Perceived Likelihood (Relative)	4.17	0.43	-.09	1.00	.10	.17	.10	-.06	.02	-.14	.13
3. Service Performance	4.12	0.70	.22	.10	1.00	.832**	-.18	-.15	-.20	-.06	.306*
4. W-O-M Intention	5.06	1.28	.14	.17	.832**	1.00	-.23	-.21	-.27	-.09	.18
5. Understanding RRI	0.63	0.49	-.448**	.10	-.18	-.23	1.00	.451**	.496**	.294*	.412**
6. Understanding Basic	0.75	0.39	-.400**	-.06	-.15	-.21	.451**	1.00	.890**	.876**	.12
7. Understanding Percentage	0.73	0.45	-.381**	.02	-.20	-.27	.496**	.890**	1.00	.560**	.22
8.Understanding Frequency	0.77	0.42	-.325*	-.14	-.06	-.09	.294*	.876**	.560**	1.00	-.02
9. Subjective Numeracy Scale	4.39	0.75	-.04	.13	.306*	.18	.412**	.12	.22	-.02	1.00

PANEL D Correlations between variable for Visual format

	M	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Perceived Likelihood (Absolute)	3.62	0.83	1.00	.09	.14	.22	.04	.10	.08	.10	-.10
2. Perceived Likelihood (Relative)	3.94	0.42	.09	1.00	-.06	-.15	.24	.486**	.387**	.484**	.21
3. Service Performance	4.18	0.67	.14	-.06	1.00	.676**	.15	-.20	-.23	-.13	-.09
4. W-O-M Intention	4.92	1.05	.22	-.15	.676**	1.00	.00	-.285*	-.27	-.24	-.28
5. Understanding RRI	0.52	0.50	.04	.24	.15	.00	1.00	.17	.14	.16	.21
6. Understanding Basic	0.90	0.27	.10	.486**	-.20	-.285*	.17	1.00	.907**	.864**	.27
7. Understanding Percentage	0.88	0.33	.08	.387**	-.23	-.27	.14	.907**	1.00	.572**	.18
8.Understanding Frequency	0.92	0.27	.10	.484**	-.13	-.24	.16	.864**	.572**	1.00	.321*
9. Subjective Numeracy Scale	4.36	0.94	-.10	.21	-.09	-.28	.21	.27	.18	.321*	1.00

* Correlation is significant at the .05 level (2-tailed).

** Correlation is significant at the .01 level (2-tailed).

Appendix 6. Distribution of responses for Understanding (RRI)

	Presentation Format					
Variable	Percentage		Frequency		Visual	
		n		n		n
Approximately 10%	52%	28	21%	10	38%	19
Approximately 170%	4%	2	4%	2	6%	3
Approximately 70%[a]	37%	20	63%	30	52%	26
Approximately 15%	7%	4	13%	6	4%	2

[a] Correct answer

Appendix 7. The survey

Block 1: Introduction and GDPR

Hur uppfattas kommunikation i vårdsamtal? Vi behöver din hjälp!

Välkommen!

Du kommer snart att svara på en enkät som är en del av vår kandidatuppsats på Handelshögskolan i Stockholm. Syftet med uppsatsen är att undersöka hur kommunikation kan uppfattas i vårdsamtal.

Din medverkan är avgörande för att hjälpa oss att slutföra uppsatsen och som tack donerar vi för varje svar 5 SEK till Barncancerfonden. Dina svar är helt anonyma och vi hanterar datan konfidentiellt.

Enkäten bör inte ta mer än 10-15 minuter av din tid. Vi är tacksamma för att du tar dig tid och hjälper oss att bidra med kunskap inom området!

Om du har några frågor är du välkommen att kontakta oss:
Andreas Lidin 24186@student.hhs.se
Daniel Mira 24785@student.hhs.se

Alla svar är anonyma och behandlas i enlighet med gällande dataskyddsförordning (GDPR).

Vänligen ta del av informationen nedan:

Projekt: BSc thesis in Business & Economics
År och termin: 2023, höstterminen
Ansvariga studenter för studien: Daniel Mira, BSc-student (24785@student.hhs.se) samt Andreas Lidin, BSc-student (24186@student.hhs.se)
Handledare och avdelning vid SSE: Patric Andersson, Associate Professor; Institutionen för Marknadsföring och Strategi

Handledarens e-postadress: patric.andersson@hhs.se
Typ av personuppgifter om dig som ska behandlas: initialer, kön, ålder och utbildningsnivå

Studentens projekt: Som en väsentlig del av utbildningsprogrammet vid Handelshögskolan i Stockholm gör inskrivna studenter ett individuellt examensarbete. Detta arbete bygger ibland på enkäter och intervjuer kopplade till ämnet. Deltagandet är naturligtvis helt frivilligt och denna text är avsedd att ge dig nödvändig information som kan röra ditt deltagande i studien eller intervjun. Du kan när som helst återkalla ditt samtycke och dina uppgifter kommer därefter att raderas permanent.

Sekretess: Allt som du säger eller anger i enkäten eller intervjun kommer att hållas strikt konfidentiellt och kommer endast att göras tillgängligt för handledare och kursadministration.

Säker lagring av data: All data kommer att lagras och bearbetas säkert av Handelshögskolan i Stockholm och kommer att raderas permanent när projektet är slutfört.

Inga personuppgifter kommer att publiceras: Uppsatsen som skrivits av studenterna kommer inte att innehålla någon information som kan identifiera dig som deltagare i undersökningen.

Dina rättigheter enligt GDPR: Du är välkommen att besöka https:/ /www.hhs.se/sv/om-oss/data-skydd/ för att läsa mer och få information om dina rättigheter relaterade till personuppgifter.

Samtycke: Jag har tagit del av informationen ovan och samtycker till att delta i studien

- ○ Ja. Vänligen fyll i dina initialer samt dagens datum (t.ex. AB 2023-12-24)
- ○ Nej

Block 2: Percentage Format

Du kommer nu att få läsa ett kort scenario som du ska föreställa dig är på riktigt.

Vänligen läs igenom scenariot noggrant. Efter att du har läst scenariot kommer du att få svara på ett antal frågor relaterade till situationen.

Var god notera särskilt fetmarkerad information!

Föreställ dig följande:

En person befinner sig på en vårdcentral och håller precis på att genomföra en hälsokontroll. Under besöket får personen svara på ett antal frågor om sitt allmänna hälsotillstånd. Läkaren lyssnar även på hjärtat och lungorna och till sist ber läkaren personen att lägga sig ned för att mäta blodtrycket.

Efter en kort stund tittar läkaren upp och berättar att personen verkar ha förhöjt blodtryck. För att utreda vad det beror på får personen lämna blodprov. Läkaren föreslår en ny besökstid samma vecka för att gå igenom resultaten och mäta blodtrycket vid ytterligare ett tillfälle.

Nästa besök:

Läkaren konstaterar även denna gång ett förhöjt blodtryck och berättar att det inte gick att se något ovanligt på blodproven.

Läkaren förklarar vidare att ett förhöjt blodtryck kan vara en indikator på att i framtiden ha en ökad risk för att utveckla olika hjärt- och kärlsjukdomar som hjärtinfarkt, hjärtsvikt, stroke och hjärnblödning.

Läkaren förklarar vidare att för människor med **normalt blodtryck** drabbas i genomsnitt **14%** av hjärt- och kärlsjukdom i

framtiden medan för människor med **förhöjt blodtryck** drabbas cirka **24%**.

Efter att ha delat denna information rekommenderar läkaren flera livsstilsförändringar som kan sänka blodtrycket och på så vis minska risken för dessa sjukdomar. Om det inte hjälper kan läkemedel vara ett annat alternativ.

Block 3: Frequency Format

Du kommer nu att få läsa ett kort scenario som du ska föreställa dig är på riktigt.

Vänligen läs igenom scenariot noggrant. Efter att du har läst scenariot kommer du att få svara på ett antal frågor relaterade till situationen.

Var god notera särskilt fetmarkerad information!

Föreställ dig följande:

En person befinner sig på en vårdcentral och håller precis på att genomföra en hälsokontroll. Under besöket får personen svara på ett antal frågor om sitt allmänna hälsotillstånd. Läkaren lyssnar även på hjärtat och lungorna och till sist ber läkaren personen att lägga sig ned för att mäta blodtrycket.

Efter en kort stund tittar läkaren upp och berättar att personen verkar ha förhöjt blodtryck. För att utreda vad det beror på får personen lämna blodprov. Läkaren föreslår en ny besökstid samma vecka för att gå igenom resultaten och mäta blodtrycket vid ytterligare ett tillfälle.

Nästa besök:

Läkaren konstaterar även denna gång ett förhöjt blodtryck och berättar att det inte gick att se något ovanligt på blodproven.

Läkaren förklarar vidare att ett förhöjt blodtryck kan vara en indikator på att i framtiden ha en ökad risk för att utveckla olika hjärt- och kärlsjukdomar som hjärtinfarkt, hjärtsvikt, stroke och hjärnblödning.

Läkaren förklarar vidare att för människor med **normalt blodtryck** drabbas i genomsnitt **14 av 100** av hjärt- och kärlsjukdom i framtiden medan för människor med **förhöjt blodtryck** drabbas cirka **24 av 100**.

Efter att ha delat denna information rekommenderar läkaren flera livsstilsförändringar som kan sänka blodtrycket och på så vis minska risken för dessa sjukdomar. Om det inte hjälper kan läkemedel vara ett annat alternativ.

Block 4: Visual Format

Du kommer nu att få läsa ett kort scenario som du ska föreställa dig är på riktigt.

Vänligen läs igenom scenariot noggrant. Efter att du har läst scenariot kommer du att få svara på ett antal frågor relaterade till situationen.

Var god notera särskilt fetmarkerad information!

Föreställ dig följande:

En person befinner sig på en vårdcentral och håller precis på att genomföra en hälsokontroll. Under besöket får personen svara på ett antal frågor om sitt allmänna hälsotillstånd. Läkaren lyssnar även på hjärtat och lungorna och till sist ber läkaren personen att lägga sig ned för att mäta blodtrycket.

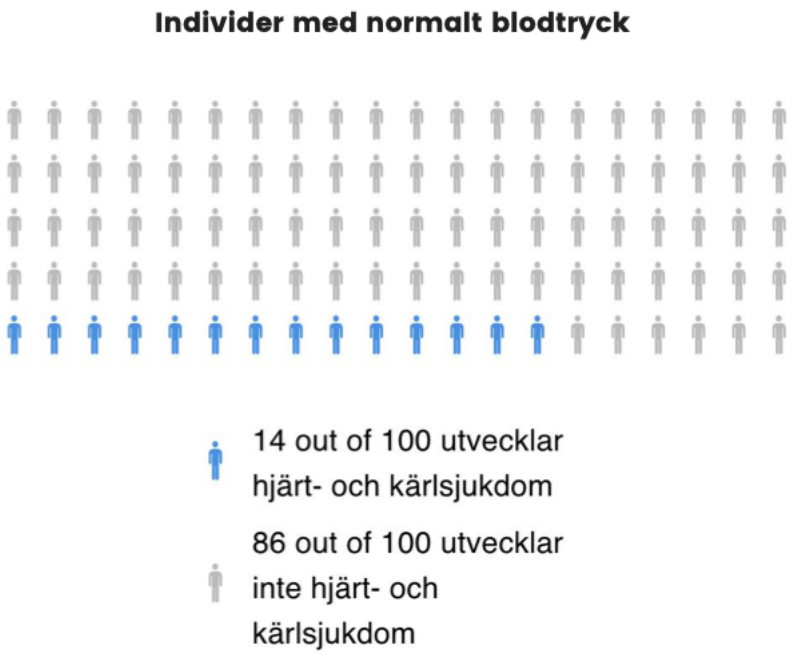
Efter en kort stund tittar läkaren upp och berättar att personen verkar ha förhöjt blodtryck. För att utreda vad det beror på får personen lämna blodprov. Läkaren föreslår en ny besökstid samma vecka för att gå igenom resultaten och mäta blodtrycket vid ytterligare ett tillfälle.

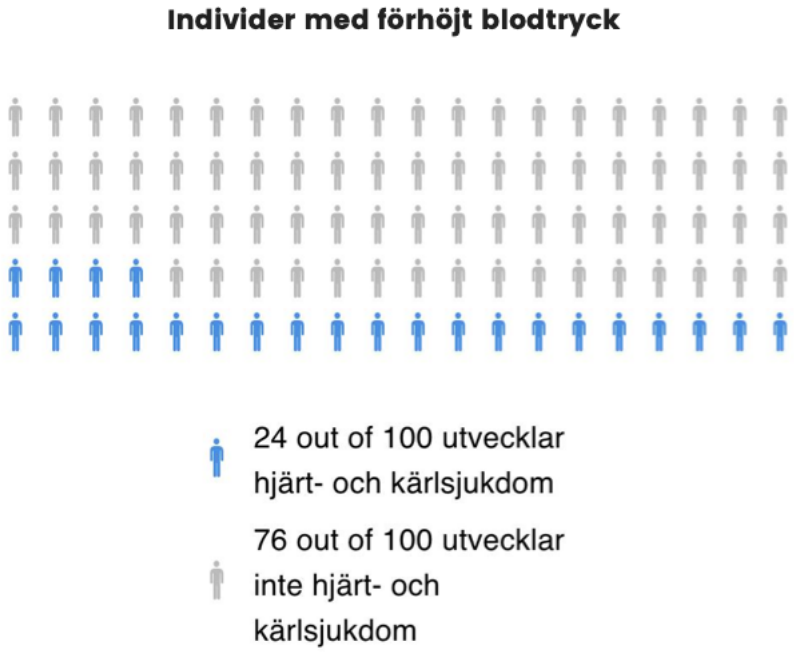
Nästa besök:

Läkaren konstaterar även denna gång ett förhöjt blodtryck och berättar att det inte gick att se något ovanligt på blodproven.

Läkaren förklarar vidare att ett förhöjt blodtryck kan vara en indikator på att i framtiden ha en ökad risk för att utveckla olika hjärt- och kärlsjukdomar som hjärtinfarkt, hjärtsvikt, stroke och hjärnblödning.

Läkaren visar även följande bild:





Efter att ha delat denna information rekommenderar läkaren flera livsstilsförändringar som kan sänka blodtrycket och på så vis minska risken för dessa sjukdomar. Om det inte hjälper kan läkemedel vara ett annat alternativ.

Block 5: Perceived Likelihood (Absolute)

Du kommer nu att få svara på ett antal frågor relaterade till scenariot du precis läste.

Det finns inget rätt eller fel – vi är enbart intresserade av din uppfattning.

Några frågor kan upplevas som repetitiva men måste ändå finnas med.

Läs igenom frågorna noggrant. Notera att de handlar om den **absoluta sannolikheten** (d.v.s. sannolikheten utan att jämföra).

Baserat på läkarens information:

	Väldigt låg	Låg	Varken låg eller hög	Hög	Väldigt hög
Om personen inte sänker blodtrycket så är sannolikheten för att utveckla hjärt- och kärlsjukdom i framtiden:	○	○	○	○	○

	Instämmer inte alls	Instämmer delvis inte	Varken instämmer eller instämmer inte	Instämmer delvis	Instämmer helt
Om personen inte sänker blodtrycket så är sannolikheten för att i framtiden utveckla hjärt- och kärlsjukdom **hög**	○	○	○	○	○

	Väldigt låg	Låg	Varken låg eller hög	Hög	Väldigt hög
Hur skulle du uppskatta sannolikheten för att utveckla hjärt- och kärlsjukdom i framtiden om personen inte sänker blodtrycket:	○	○	○	○	○

Block 6: Perceived Likelihood (Comparative)

Notera att följande frågor är **jämförande** (d.v.s. jämfört med normalt blodtryck)**.**

Baserat på läkarens information:

	Mycket mindre	Mindre	Varken mindre eller högre	Högre	Mycket högre
Om personen inte sänker blodtrycket så är sannolikheten för att utveckla hjärt- och kärlsjukdom i framtiden:	○	○	○	○	○

	Instämmer inte alls	Instämmer delvis inte	Varken instämmer eller instämmer inte	Instämmer delvis	Instämmer helt
Om personen inte sänker blodtrycket är det **mindre** sannolikt att utveckla hjärt- och kärlsjukdom i framtiden:	○	○	○	○	○

Block 7: SERVPERF

Nedanstående påståenden är relaterade till scenariot. Försök att svara även om något verkar otydligt eller svårt! **Om du är osäker – gå på din magkänsla.**

Läs igenom påståendena noggrant.

Svara baserat på **scenariot** du precis läste (försök att inte tänka "generellt").

Hur väl instämmer du med följande påståenden:

	Instämmer inte alls	Instämmer delvis inte	Varken instämmer eller instämmer inte	Instämmer delvis	Instämmer helt
Det går att lita på läkarens förmåga att hantera patientens hälsoproblem.	○	○	○	○	○
Läkaren visar en vilja att hjälpa patienten.	○	○	○	○	○
Läkaren är förtroendeingivande.	○	○	○	○	○
Läkaren har tillräcklig kunskap för att kunna besvara eventuella frågor som patienten har.	○	○	○	○	○
Läkaren vill patienten väl.	○	○	○	○	○
Läkaren förstår patientens behov.	○	○	○	○	○
Det är viktigt att vara uppmärksam när du genomför enkäten. Tryck på svarsalternativet "Instämmer inte alls".	○	○	○	○	○

Block 8: Word-of-Mouth Intention

Är det troligt att du skulle rekommendera den här läkaren till en vän eller kollega?

Nej, absolut inte	Nej, i stort sett inte	Nej, troligen inte	Tveksam	Ja, troligen	Ja, i stort sett	Ja, absolut
○	○	○	○	○	○	○

Block 9: Understanding (percentage)

Nu är mer än hälften av enkäten klar!

Nu kommer en fråga som handlar om riskerna som presenterades i scenariot. För att påminna dig om vad läkaren förklarade:

"Läkaren förklarar vidare att för människor med **normalt blodtryck** drabbas i genomsnitt **14%** av hjärt- och kärlsjukdom i framtiden medan för människor med **förhöjt blodtryck** drabbas cirka **24%**."

Med hur mycket **ökar** risken för en person med förhöjt blodtryck jämfört med en person med normalt blodtryck? Den procentuella ökningen är:

- cirka 10%
- cirka 170%
- cirka 70%
- cirka 15%

Av **1000 människor** med förhöjt blodtryck, hur många förväntas utveckla hjärt- och kärlsjukdom i framtiden?

Ange antal (st.)

Av **1000 människor** med förhöjt blodtryck, hur många procent förväntas utveckla hjärt- och kärlsjukdom i framtiden?

Ange svar i % – endast siffra.

Om du vill utveckla dina svar eller har en kommentar, vänligen skriv nedan:

Block 10: Understanding (frequency)

Nu är mer än hälften av enkäten klar!

Nu kommer en fråga som handlar om riskerna som presenterades i scenariot. För att påminna dig om vad läkaren förklarade:

"Läkaren förklarar vidare att för människor med **normalt blodtryck** drabbas i genomsnitt **14 av 100** av hjärt- och kärlsjukdom i framtiden medan för människor med **förhöjt blodtryck** drabbas cirka **24 av 100**. "

Med hur mycket **ökar** risken för en person med förhöjt blodtryck jämfört med en person med normalt blodtryck? Den procentuella ökningen är:

- cirka 10%
- cirka 170%
- cirka 70%
- cirka 15%

Av **1000 människor** med förhöjt blodtryck, hur många förväntas utveckla hjärt- och kärlsjukdom i framtiden?

Ange antal (st.)

Block 11: Understanding (visual)

Nu är mer än hälften av enkäten klar!

Nu kommer en fråga som handlar om riskerna som presenterades i scenariot. För att påminna dig om vad läkaren förklarade:

Individer med normalt blodtryck

14 out of 100 utvecklar hjärt- och kärlsjukdom

86 out of 100 utvecklar inte hjärt- och kärlsjukdom

Individer med förhöjt blodtryck

24 out of 100 utvecklar hjärt- och kärlsjukdom

76 out of 100 utvecklar inte hjärt- och kärlsjukdom

Med hur mycket **ökar** risken för en person med förhöjt blodtryck jämfört med en person med normalt blodtryck? Den procentuella ökningen är:

- ○ cirka 10%
- ○ cirka 170%
- ○ cirka 70%
- ○ cirka 15%

Av **1000 människor** med förhöjt blodtryck, hur många förväntas utveckla hjärt- och kärlsjukdom i framtiden?

Ange antal (st.)

Av **1000 människor** med förhöjt blodtryck, hur många procent förväntas utveckla hjärt- och kärlsjukdom i framtiden?

Ange svar i % – endast siffra.

Om du vill utveckla dina svar eller har en kommentar, vänligen skriv nedan:

Block 12: Numeracy

Tack! Nu är du färdig med frågorna relaterade till scenariot.

Vänligen observera att scenariot och läkarens rekommendationer är påhittade och endast används av illustrativa skäl. Riskerna förknippade med högt blodtryck är baserat på data för unga människor men har förenklats väsentligt.

Här kommer några avslutande allmänna frågor:

	Mycket dålig	Dålig	Varken dålig eller bra	Ganska bra	Bra	Mycket bra
Hur bra är du på att hantera bråktal?	○	○	○	○	○	○
Hur bra är du på att hantera procent?	○	○	○	○	○	○
Hur bra är du på att räkna ut 15% i dricks?	○	○	○	○	○	○
Hur bra är du på att räkna ut hur mycket en skjorta kommer att kosta om den har 25% rabatt?	○	○	○	○	○	○

	Mycket lite	Lite	I någon mån	I viss grad	I hög grad	I mycket hög grad
När du läser nyheter, hur användbara tycker du att tabeller och grafer är som en del av artikeln?	○	○	○	○	○	○

	Föredrar alltid ord 1.	2.	3.	4.	5.	Föredrar alltid siffror 6.
När människor berättar om sannolikheten för att något ska hända, föredrar du att de använder ord (som "det händer sällan") eller siffror (som "det är 1% chans")?	○	○	○	○	○	○

	Föredrar alltid procent 1.	2	3	4	5	Föredrar alltid ord 6.
När du hör en väderprognos, föredrar du bedömningar som använder procent (t.ex. "det är 20% sannolikhet för regn idag") eller bedömningar som endast använder ord (t.ex. "det är en liten sannolikhet för regn idag")?	○	○	○	○	○	○

	Aldrig	Sällan	Ibland	Ganska ofta	Ofta	Mycket ofta
Hur ofta tycker du att siffror och data är användbara?	○	○	○	○	○	○

Block 13: Demographic variables

Nu är det inte långt kvar. Vänligen svara på följande frågor om dig själv:

Vänligen ange kön:

Man ○ Kvinna ○ Icke-binär ○ Annat ○ Vill ej ange ○

Ange ålder (svara med en siffra, t.ex. 25):

Vilken är din högsta avslutade utbildning?

Grundskola eller motsvarande	Gymnasieexamen	Yrkesutbildning	Högskola eller universitet (1–3 år)	Högskola eller universitet (>3 år)
○	○	○	○	○

Block 14: Survey Evaluation

Slutligen vill vi gärna veta vad du tycker om vår enkätundersökning. Vänligen besvara sista frågorna nedan:

	Instämmer inte alls	Instämmer delvis inte	Varken instämmer eller instämmer inte	Instämmer delvis	Instämmer helt
Frågorna var tydligt formulerade	○	○	○	○	○
Svarsalternativen var tydligt formulerade	○	○	○	○	○
Scenariot var realistiskt	○	○	○	○	○
Frågorna försökte påverka dina svar på något sätt	○	○	○	○	○

Läkaren i scenariot förklarade riskerna associerade med:

○ Diabetes ○ Förhöjt blodtryck ○ Cancer ○ Njursten

Appendix 8. Disclosure of AI-tools

Based on the current status of generative AI-tools and their availability, SSE requires an AI-disclosure (as of December 2023).

The authors of this thesis verify that no AI-tools have been used during the writing of this thesis or during the collection of data. Our understanding is that the current generation of generative AI (such as Chat GPT-3.5/4) is not sophisticated enough to provide reliable and unbiased output. To our knowledge, tools such as Chat GPT have limited access to information published after a certain cut-off date and have been reported to provide users with references to articles that do not exist.

However, we do recognize that moving forward AI-tools might prove useful in certain stages of the thesis writing process. If the current generation of AI was able to retrieve published articles in real-time, the literature search could be more efficient. There could also be a use-case in terms of generating visual content for experiments based on prompts from the researcher.

As of today, we do not consider the application of AI to be viable in thesis writing based on the reasons stated above.