Erasmus University Rotterdam
Erasmus School of Economics
Master Thesis M.Sc. Behavioural Economics

At Risk of Endometriosis: Is Receiving the Right Treatment Enough?

Stephan Rothenberger (646649)

Supervisor:	dr. Johan de Jong
Second assessor:	prof. dr. Peter Wakker
Date final version:	24th July 2023

Abstract

Endometriosis is a severely underdiagnosed disease that affects more than 190 million women worldwide, partially heavily decreasing their quality of life. We combine this problem with the fact that the preference for pure uncertainty resolution, i.e., seeking information at a cost without direct benefit, is not well understood in healthcare. To create first evidence of this preference, we leverage the unique situation of diagnostic pathway choices that patients with Endometriosis risk face, thereby also improving the understanding of this strongly underdiagnosed disease. We elicit the stated preferences of women in a hypothetical decision scenario with exogenous between-subject varied symptom levels in a discrete choice experiment. We vary the accuracy of treatment outcomes and the cost of diagnostic certainty to display different realistic trade-offs with which we measure subject preferences. By collecting additional information on health status, attitudes, and choice reasoning we differentiate the role of uncertainty resolution from other reasons. We find a substantial deviation from the behavior predicted by neoclassical theories, which indicates a preference for uncertainty resolution. We do not find information avoidance or the symptom level to drive the diagnostic pathway choice in our experiment.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

When hearing about a severely underdiagnosed disease with 10 years of diagnostic delay and limited treatment options one would imagine a rare disease that is unlikely to affect one's life. However, Endometriosis (EMS) affects an estimated 8% of women of childbearing age and we estimate a yearly cost of over 100 million euros for the German healthcare market alone. The disease, simply described as endometrium-like tissue outside the uterus, which can cause symptoms such as scarring, severe pain, and infertility, impacts over 190 million women worldwide every day, thereby impacting their quality of life, e.g., through their decreased ability to work and to lead a regular social life (ESHRE GDG, 2022).

Recently, a new diagnostic test with very high accuracy was introduced to the market (Bendifallah, Dabi et al., 2022; Bendifallah, Suisse et al., 2022). This test represents an alternative to the currently recommended diagnosis through a specialist practitioner who uses imaging techniques like ultrasound. The imaging techniques can be as accurate as the test if the specialist is well educated and equipped with the latest technology (ESHRE GDG, 2022). Both enable accurately pinpointing the indication (simply said, a reason to recommend a certain treatment), which is sufficient for treating the patient accurately. Yet, only the diagnostic test allows for the factual identification of EMS. In this case, the patient has no benefit in terms of treatment outcomes by choosing the diagnostic test over the specialist's diagnosis but has the downside of the test cost. Her key benefit is therefore that she factually knows the diagnosis and thus, e.g., can more easily gain social acceptance. Assuming that there are no other reasons to choose the diagnostic test that add value in the traditional sense, this can be defined as a preference for Uncertainty Resolution (UR).

More generally said, a pure preference for UR can be defined as a preference for the resolution of information (at a cost) that does not lead to additional utility. I.e., resolving an uncertainty without a direct benefit, such as cost savings or improved treatment outcomes. In the described case we find a rare real-world decision situation that depicts the possibility of uncertainty resolution (UR).

By merging these two topics, researching the choices of individuals with EMS risk has double relevance. For one, it can possibly help to reduce the delay in treatment, the healthcare cost, and the quality of life impact mentioned above. For another, it can increase the understanding of UR behavior in healthcare. While there is research on UR it mainly focuses on financial behavior, lotteries, and timing of information, e.g., the work of Chew and Ho (1994) and Lovallo and Kahneman (2000) among many others. We are not aware of research focusing on UR in the medical context. Lillrank (2003) shows qualitatively that UR plays a role in the decision-making of women with back pain but does not separate it from actual treatment outcomes.

So far healthcare research has focused on understanding why people actively avoid information and its consequences, with little to no attention given to the opposite case. Karlsson et al. (2009) first show that hedonic motivations can incentivize information-avoiding behaviors. This opens the question of whether information-seeking behavior could also be purely hedonic. One might ask: do people, at a cost, seek information that does not create direct personal or economic value? If so, why do they do it, i.e., wherein does their utility of doing so lie, and which consequences or implications does this have for patient behavior, design and cost of healthcare systems, and patient choice architecture?

To make the first step in closing this research gap, we aim to identify if this seeking of information that does not add direct value (i.e., UR) is, in fact, a key choice driver in patients' decision-making, and which behavioral drivers of medical information-seeking it might correlate with. Concisely, we ask:

Research Question:
Is a correct treatment enough for patients with endometriosis symptoms, or do they also seek certainty with their choice of the diagnostic pathway?

To answer this question we leveraged a Discrete Choice Experiment (DCE) wherein online participants received the needed information about EMS to make educated decisions and then answer ten trade-off questions selecting between different diagnostic pathways options. The forced choice trade-offs always included two options of the following three labeled alternatives: no diagnosis, the diagnostic test, and the specialist practitioner. The diagnostic test varied in price and the specialist practitioner varied in treatment accuracy, thereby displaying different real-world choice options. Trading between these realistic choice options, individuals stated their preferences and thereafter shared their EMS-related health status, their healthcare attitudes, and reasons to choose the diagnostic test given no improved treatment accuracy. In this process, participants got randomly allocated to a less or more severe symptom description based on which they were asked to make all their decisions.

Based on the answers from 300 women, our main analysis implements a panel mixed logit model fit for individual-specific choices of the labeled alternatives. Thereby we find a substantial deviation from the neoclassical economic prediction, as participants do not only demonstrate a willingness to pay for increased treatment accuracy but also for the resolution of uncertainty, without direct benefit to economic or health outcomes. While we cannot clearly isolate UR from other minor differences between the diagnostic test and the specialist practitioner (e.g., disliking ultrasounds), we show that individuals do not name reasons apart from UR for this preference and that a substantial amount explicitly self-reports their preference for UR. Further, we do not find information avoidance to drive the diagnostic pathway choices in our setting. Lastly, we did not find the hypothetical symptom level to play a role in decision-making. These findings acknowledge the importance of UR in healthcare settings, thereby calling for future research, and can help understand the decision-making of patients with EMS risk better and thus possibly inform the design of healthcare systems, legislation, and choice architecture implemented by doctors every day.

In chapter 2 we lay out how patient decisions are seen in classical economic theory and how these classical theories can and have already been extended by behavioral aspects such as information avoidance, leading up to why we hypothesize uncertainty resolution might be a key decision driver. In chapter 3 we detail our choice of a Discrete Choice experiment and how it is constructed and analyzed, followed by the results and discussion, including the conclusion, in chapter 4 and chapter 5 respectively.

Chapter 2

Literature Review

In the following, we will give a short introduction to EMS, followed by the introduction of various more traditional concepts as the starting point of our research. We then explore relevant concepts of information-seeking behavior that complete these or put these models into question. We will merge the individual concepts with prior mentioned ones and add information on healthcare practice and EMS where needed, thereby leading toward the individual hypotheses.

2.1 Endometriosis

The term Endometriosis (EMS) describes a recognized chronic inflammatory disease. It is defined by endometrium-like tissue in areas that are not the uterus, mainly in the abdominal cavity, the pelvis, and the genital organs. The development of this tissue is estrogen-dependent, thus mainly affecting women of childbearing age. The growth of this tissue often ensues severe pain, inflammation, scarring, hormone level fluctuation, digestive problems, tiredness and fatigue, urinary tract disease, infertility, and many more symptoms impacting the patient's life and the people around them. Estimates suspect that 6-10% of women of childbearing age suffer from EMS - more than 190 million women worldwide. The estimates vary rather strongly as EMS is severely underdiagnosed. There is a very substantial delay in diagnosis of 8-12 years. This makes EMS stand out from diseases like Crohn's disease, type 2 diabetes, and rheumatoid arthritis that have similar impact on health, quality of life, society, and economy (Adamson et al., 2010; ESHRE GDG, 2022; Pugsley & Ballard, 2007).

Underdiagnosis Some causes for underdiagnosis are the diagnostic complexity, the unclear aetiology (simply said the causes and how they work), the normalization of symptoms, and a lack of awareness on the provider and patient side (Hudson, 2021; Illum et al., 2022). It is important to acknowledge that a major part of the diagnostic delay is caused by primary care practitioners who fail to acknowledge the relevance of symptoms and thus do not refer patients to a specialist until an average of 7 visits, as a study by Nnoaham et al. (2011) over 10 countries found.

Impact and relevance Due to the lack of diagnosis and attention for EMS in the past, the picture of its impact on health and society is not very clear. It is out of question, that EMS can heavily impact the overall quality of life. It impacts well-being to an extent preventing women around the world from going to school, working, or participating in other activities. Further, it affects sexual and reproductive health (Berkley et al., 2005; WHO, 2023).

In one intent to quantify these consequences, Bianconi et al. (2007) estimate the cost to be 2,773.80 euros per patient in Italy, adding to a total sum of 54 million euros per year. As the is no estimate for Germany, we create a rough estimate based on these numbers. We build on the estimation for Italy while considering the higher population and the change in buying power since 2007 but do not account for the higher healthcare expenditure in Germany. This sums up to about 3,727 euros per patient or over 102 million euros per year in cost for the German Healthcare system. Independent of this, analyzing Medicaid data in the USA Soliman et al. (2019) show that the cost is not only created through the presence of symptoms, which also non-EMS patients may have. They found healthcare costs to be more than double for women with EMS compared to women without EMS having the same symptoms. Nnoaham et al. (2011) found an average loss of 10.8 productive work hours per week, mainly due to reduced efficiency. This also has substantial implications for women suffering from EMS in their ability to compete and be fairly assessed in the job market. They also find that EMS patients have a higher quality of life loss than non-EMS patients with the same symptoms. These monetary and non-monetary costs highlight the importance of (early) diagnosis.

Diagnosis For health and consequentially the mentioned economic costs early diagnosis and effective treatment are key. However, this is often not achieved (WHO, 2023). A study on the accuracy of diagnosis through video laparoscopy found major disagreements between 108 specialists (Kahneman et al., 2021). While, until recently, the gold standard included laparoscopic surgery in combination with visualization and histological confirmation, the newest guidelines do not recommend laparoscopy for diagnosis of the vast majority of patients, thereby avoiding surgery (ESHRE GDG, 2022; Goehring et al., 2022). The latest gold standard for diagnosis includes clinical examination and imaging (ultrasound or magnetic resonance imaging) and explicitly opposes the use of biomarkers in endometrial tissue, blood, and menstrual or uterine fluids (ESHRE GDG, 2022).

Further, there is a novel diagnostic test with limited availability around the world. It is currently (July 2023) accessible in Germany at a price point of 800 euros. The test works similarly to a Covid-19 saliva PCR test, collecting saliva with a swab in which one can identify chemical biomarkers (in the form of proteins, differing from the above-mentioned biomarkers) using microRNA technology. The test benefits from a very high sensitivity of 97%, a specificity of 100%, and a diagnostic accuracy of 98% (Bendifallah, Dabi et al., 2022; Bendifallah, Suisse et al., 2022; Eluthia, 2023).

While both methods, the imaging technology used by a specialist and the diagnostic swab test, are non-invasive and do not bring a significant amount of discomfort, the key difference between the two methods is that both can clearly identify the indication (the reason a treatment would be used, e.g. a cancer tumor for cancer, cystic lesions for EMS) but only the diagnostic test can factually confirm the underlying disease, while the imaging technique remains with a suspicion of the actual underlying cause of the treated indication, which we refer to as a suspicion diagnosis. While both methods are highly accurate in identifying the indication and hence can offer a highly accurate treatment, the accuracy of the suspicion diagnosis is subject to many factors such as the practitioners skill level and technological equipment.

Treatment The treatment following the diagnosis is constantly developing, with major shifts of the prescribed medication in Germany (Goehring et al., 2022). However, the shift has stayed limited to the realm of treating or suppressing symptoms. Typical treatment includes painkillers, hormone treatment such as combined hormonal contraceptives, progestogens (a type of "pregnancy hormone"), hormone agonist and antagonists, and aromatase inhibitors (an enzyme involved in estrogen production). Surgery, having been replaced by imaging techniques as a diagnostic tool, is now also merely recommended to be offered to patients but is not a standard part of treatment, as the positive impact of laparoscopic surgery on quality of life and overall pain is not sufficiently proven and thus does not allow for a valid conclusion. To date there is no cure (ESHRE GDG, 2022).

Concluding, it is important to keep in mind the complexity and ambiguity of symptoms, the often vast delay of diagnoses, the limited treatment, and the substantial impact on women, the healthcare system, and society as a whole in the following review of economic and behavioral subject matters, as well as in the complete experimental design and analysis as they reason design choices, research motivation, and ultimately the patient's choice and therefore the study results.

2.2 Economic Theory

When creating hypotheses on behavior it is common practice to use traditional neoclassical models as a basis and then argue and analyze where and why behavior may differ (Golman et al., 2017; Li et al., 2021). Economic models addressing how information drives choice typically predict behavior singularly on the basis of direct economic or personal benefit (utility) and build on the assumption that individuals cannot act on a situation (e.g. treat a disease) without revealing (in-)complete information (e.g., a diagnosis). For example Li et al. (2021) apply this, testing the effect of price on the choice of diagnostic tests for cancer. Following neoclassical models, a person should be willing to pay more for the diagnostic test if they are at a higher risk of cancer, ceteris paribus, given situations where the specialist practitioner offers less treatment accuracy. Further, it is assumed that a cancer treatment cannot be initiated without the treatment. Hence, one would also expect a monotonic increase in uptake depending on individual risk for cancer, ceteris paribus.

In the case of EMS, this means that patients and the healthcare system only consider the cost and benefit of a diagnosis and the consequential treatment. They do not consider indirect effects such as peace of mind, anxiety, social acceptance, or just wanting to know as part of the costs or benefits and thus utility function. Further, they can only act on a diagnostic result and cannot treat without a diagnosis. Simply said the patient's utility is equal to the health benefit minus the cost.

Price As one may expect, experiments on diagnostic test uptake confirm the expected economic theory. E.g., Li et al. (2021) show a 50% decline in test uptake with a tripling of the price for diabetes tests. The neoclassical economical model would imply that with an increase in price, high-risk and wealthier subjects remain, while the others self-select not to do the test.

In the choice setting of this experiment, which is explained to detail in section 3.1, visiting a specialist practitioner does not come at a cost. However, the diagnostic test is offered at different cost levels, depending on the co-pay (share not covered by insurance). This mirrors the German healthcare market.

As the difference of the diagnostic methods does not change the treatment outcome, given the specialist practitioner does not lack knowledge or technical equipment, the neoclassical economical model would predict everyone to seek a diagnosis with the specialist, as long as their subjective risk is high enough to justify the "cost" of investing time, etc. for the appointment. No subject should choose to pay for the diagnostic test if the specialist does not lack knowledge or equipment. Hence we hypothesize:

H1: A majority of individuals make the cost-minimizing, benefit-maximizing choice of going to an endometriosis specialist if there is no improvement in treatment outcome.

H2: The diagnostic test is chosen more often with decreasing price.

Accuracy of treatment Taber et al. (2015) identify that various reasons related to trust in the medical professional's ability contribute to driving people away from the doctor. While the diagnostic accuracy and the accuracy of treatment are under-researched topics, the available research from varying indications makes this lack of trust, which is often built on the personal experiences of patients, reasonable.

Available data typically focus on the accuracy of diagnosis of specific diseases. While this is by no means representative of EMS diagnosis, one can understand the strong variation in diagnostic accuracy, which thereby impacts the accuracy of treatment. On the lower end of diagnostic accuracy, research finds the identification skill of general practitioners with Mitchell et al. (2011) finding around 74% sensitivity for dementia and 63% for cognitive impairment, and Mant et al. (2007) finding around 80% sensitivity for atrial fibrillation. Garrido et al. (2020) show that there is a substantial difference in skill level between general practitioners and specialists, with only 35% of Brazilian general practitioners but 96% of Brazilian specialists knowing the key rule for skin cancer. Hirst and Smith (2020) find that specialist optometrists are accurate in 98.6% of cases of a pterygium (also known as "surfer's eye"). This small overview shows that patients have to face a substantial variation in diagnostic and therefore treatment accuracy.

As knowledge of the actual diagnostic accuracy for the vast majority of indications is lacking, the true accuracy is often hidden. Thus, the beliefs of patients build on reviews or other factors that influence their own assessment of to what extent they can trust a doctor's expertise. Bista et al. (2021) find that healthcare avoidance is strongly correlated with a lack of trust in doctors. We remove this variation in personal assessment by supplying a figure for the accuracy of treatment in our model in each trade-off option, as explained in detail in chapter 3. In this experiment accuracy of treatment is defined as the treatment being the correct choice to treat the indication, independent of the actual underlying disease.

Level of symptoms Neoclassical models would predict increased uptake with increased symptoms or risk as the utility gain in treatment would be larger. However, the findings by Li et al. (2021), do not entirely support this. They find that, with increasing price, both low and high subjective risk individuals select out of the test. In the medical context, subjective risk can be seen as the risk one personally sees for themselves, e.g. their subjectively assumed probability of having cancer. This assumption may or may not be based on facts or knowledge. While they do show an increased uptake rate for patients with a higher subjective risk of diabetes when the test is free, they do not find the same pattern when looking at diagnostic tests for cancers. They hypothesize, that the severity of cancer and lower success probability of treatment may drive people away. To eliminate variability through subjective risk perception, we share a factor that estimates the actual, objective risk with the study subjects as further discussed in section 3.3.

Ballard et al. (2008) show that with many symptoms the odds ratio, i.e. the objective (factual) risk of having EMS, dramatically increases. Also, as discussed above, women with EMS suffer from a higher quality of life impact through symptoms than women with the same symptoms that do not have EMS (Nnoaham et al., 2011). These two facts connect the level of experienced symptoms directly with the subjective risk studied by Li et al. (2021). While EMS does not come with the risk of death like cancer, treatment is nonetheless very limited. Consequently, we hypothesize that the symptom level does not impact the choice of diagnostic method and therefore the test uptake (which is only available at a cost).

H3: Increasing symptoms do not impact the choice of diagnostic method, and therefore the uptake in diagnostic tests.

2.3 Information Avoidance

Golman et al. (2017) describe active information avoidance as situations in which the free availability of information is known to an individual who, nonetheless, chooses not to obtain or even to avoid the information. They show that the topic is widely researched in different fields from management and financial behavior to political polarization.

However, this behavior is often driven by some kind of utility, e.g., only the onset of symptoms or positive results of a diagnostic test for a disease may force the reality upon a person that to date is living a happy life. In fact, the concave form of a utility function by itself implies that, in the absence of material consequences, information should be avoided, as the negative effect of not reaching expectations is larger than the opposing effect (Golman et al., 2017).

Kőszegi (2003) formalizes this behavior into a model, wherein negative beliefs about the future enter the patient's utility function. He thereby opens up the question of the role which healthcare professionals, government policies, and health education play in achieving optimal health outcomes through impacting information-seeking behavior and argues, that these "responding" parties must take the anticipatory emotions into account (Kőszegi, 2004, 2006; Kőszegi & Rabin, 2008).

Schweizer and Szech (2018) offer a more granular look at what degree of information revelation is optimal for life-changing information, like hereditary diseases (some research suggests that EMS may be hereditary to an extent (ESHRE GDG, 2022)). They suggest that providing a test with an uncertain positive (disease confirming) but a certain negative outcome is often optimal, as a test with a certain positive outcome can be scary and may lead to decreased uptake in testing. While this model seems plausible in the case of diseases such as preventative cancer or diabetes diagnoses, the situation and therefore choice structure many patients with (hereditary) risk for EMS are in differs strongly. Considering the often significant delay in diagnosis, and that symptoms set in far prior to a diagnosis in the vast majority of cases (ESHRE GDG, 2022), it seems unlikely that patients wish to remain with partial uncertainty in case of a positive diagnosis. There seems to be no benefit in being able to ignore or question the accuracy of the diagnosis, as the reality of symptoms is already very much present in their daily lives.

Ostrich Effect The term ostrich effect was first used for describing a specific behavioral pattern of information acquiring and attending in finance by Galai and Sade (2003). They described it as the avoidance of apparently risky situations, by metaphorically sticking one's head into the sand, i.e. pretending the situation does not exist. It is key to point out, that unlike in other situations of information avoidance where there may be no information whatsoever (hence a neutral situation), here there is always an initial signal, based on which the reaction of the individual is based, e.g. ignoring an unusual feeling tingling in ones fingers which could indicate various problems from carpal tunnel syndrome to multiple sclerosis, because the tingling is a first negative signal of a quite possibly negative outcome.

Karlsson et al. (2009) first constructed a model wherein the utility from beliefs and information does not only add into the utility function as in the above-described behavioral model by Kőszegi (2003) but where the utility may incentivize actually controlling or regulating said beliefs out of hedonic reasons. In the case of EMS this can be seen in the degree to which individuals seek additional information after "receiving" negative "news", i.e. experiencing symptoms.

Evidence of the Ostrich Effect in healthcare is laid out by Panidi (2015). She brings the two effects of loss aversion and information aversion together by looking at data for hypertension, diabetes, and chronic lung disease in the Netherlands and showing that there is a negative correlation between preventive testing and loss aversion. Again, given the situation of many individuals who receive (experience) information (symptoms) that may indicate EMS, often including substantial pain from an unknown source and a lack of understanding from others, one might question if the ostrich effect plays a relevant role in the diagnostic choices of individuals with risk of EMS.

H4: The ostrich effect, which refers to the tendency to avoid negative information, is not a relevant factor in the decision between diagnostic options for EMS, contrary to most indications.

2.4 Decision drivers in medical care

The reasons patients avoid information in healthcare settings are manifold. In a qualitative analysis Taber et al. (2015) categorize general themes. Next to traditional reasons like cost and lack of health insurance (54%), which are not a constraint given the design of the experiment (doctors are available and free of cost, hence availability or cost can not lead to information avoidance), patients name discomfort (27%) and a low perceived need to seek medical care (12%) as meta reasons to avoid care.

Low perceived need looking at patients of a British breast cancer clinic, Nosarti et al. (2000) find that 33.4% of women did not think their symptoms were serious. Taber et al. (2015) identify multiple sub-dimensions of low perceived need, among the largest being the belief that the illness will improve over time, patients not feeling sick enough, or not experiencing the impact of an illness, i.e. not feeling sick. Further, some try to take care of the illness themselves.

Fear and Anxiety In qualitative research by Taber et al. (2015) 26% name fearing a serious illness as the reason why they avoid doctors visits. Similarly, Nosarti et al. (2000) identify being scared of the diagnostic result as a main reason to delay a doctor's visit in breast cancer patients, wherein the effect increased with increased symptoms. Worries about having cancer were also shown to negatively predict entry into a counseling project (Bowen et al., 1999). Further, Kash et al. (1992) show a correlation between anxiety and not following regular self-examination as well as not attending clinical examinations.

The above-described dynamics of information avoidance behavior indicate that high levels of anxiety, low perceived need, and discomfort with doctor visits may cause information avoidance.

H5: If subjects do choose not to get diagnosed, their choice is highly correlated with factors of information avoidance in terms of, anxiety, discomfort, or low perceived need of going to the doctor, i.e., avoiding a diagnosis for one or multiple of these reasons.

2.5 Uncertainty Resolution

Golman et al. (2017) point out that information is often costly to obtain, giving the example of a costly medical test, wherein not taking a test may indeed stem from the preference of avoiding information but may just as much originate from the cost, or a mix of the two factors. However, if individuals seek information, that does not add value in the strict economical sense, although there is a cost, they clearly show a preference for revealing information, here referred to as *Uncertainty Resolution.*

To date uncertainty resolution has mostly been researched in the context of time, especially using lotteries. For Example, Lovallo and Kahneman (2000) find a clear preference for earlier resolution of uncertainty when the expected value of a lottery is lower. Opposingly, there is a preference for delaying the resolution of uncertainty when there is an expectation, and thereby connected enjoyment of hoping for, a positive outcome (Chew & Ho, 1994). While there is no direct economic value in uncovering the uncertainty of a lottery earlier or later, in these cases uncertainty must be uncovered at the point of payment, when the lottery win or loss gets executed. In healthcare, this is not necessarily the case. The EMS treatment can be accurate without a factually secured diagnosis of the disease, simply relying on the correct identification of the indication. Hence, one does not pay for earlier or later resolution of uncertainty that does not add economic value but pays for uncertainty resolution in itself (in the form of the diagnostic test) that does not add economic value.

In healthcare, information avoidance has been a major topic of interest to economists due to its personal and economic consequences (Golman et al., 2017), there while uncertainty resolution has been overseen. this is possibly due to the above-described dynamic, that resolving uncertainty that does not add direct value seems irrational and, one might think, does not directly impact economic or personal outcomes. However, understanding if and why a preference for uncertainty resolution does happen can help design healthcare systems and information architecture, and model patient preferences. E.g., if patients experience anxiety through unresolved uncertainty, resolving the uncertainty can have a positive effect on their health, and in a further step reduce potential costs for the healthcare system.

The core reason why in some cases, including EMS diagnosis, the typical patterns of information avoidance and the ostrich effect may not hold seems to lie in the nature of Myopic Loss Aversion. Myopic Loss Aversion is defined as the combination of loss aversion and the tendency to frequent information updating using mental accounting (Benartzi & Thaler, 1995), meaning individuals are more sensitive to decreasing than to increasing levels of health, as repeated frequent negative information leads to higher perceived losses than repeated frequent positive information leads to perceived gains. This concept, with the implied notion of being inattentive to, hiding from, or avoiding news expected to be unfavorable, is a fundamental part of the above-described theories (Galai & Sade, 2003; Golman et al., 2017; Karlsson et al., 2009).

It seems almost certain, that Myopic Loss Aversion cannot be applied in the same way for people suffering from symptoms that may originate from EMS as these individuals frequently and involuntarily receive negative updates about their situation in the form of unpleasant symptoms. Further, deciding to seek a diagnostic result for EMS does not significantly impact an individual's utility, as the negative news in the case of a positive test merely gives an explanation. The prospect of having to live with EMS (as there is no cure), when already suffering from symptoms can be seen as significantly less impactful than, e.g., a cancer diagnosis that comes with the risk of death, significant changes to one's lifestyle, and new uncertainties about the future.

Indeed, Li et al. (2021) show that information avoidance varies over disease type, due to factors such as severity and perceived control over the disease. Hence, while the decision situations may seem similar, observing the situation that many women potentially suffering from EMS are in may imply the opposite behavior. In a situation with an initial negative signal (as given in the ostrich effect) hedonic information seeking (seeking information that does not add direct value), in contrast to the usual hedonic information avoidance for negative information, may be the case. Research on this opposite effect is extremely limited. Lillrank (2003) shows that women seek uncertainty resolution in back pain treatments to gain social validation but also to receive the correct treatment, hence not being purely driven by hedonic motivations.

As is the case with information avoidance, one may take different stances if such behavior is in line with economic theory, given a case where the information does not add direct economic or personal value in the sense of a cure or better treatment, but solely a potential hedonic value, e.g., through social recognition, acceptance or simply the fact that one can now be sure about an issue (hence, Uncertainty Resolution in the strict sense). We hypothesize that this hedonic value drives people to choose the diagnostic test, even in cases where there is no benefit in the treatment outcome compared to a suspicion diagnosis through a doctor. If this hypothesis holds true we make a strong case for uncertainty resolution in healthcare settings, which consequently should be included in utility functions of patient decisions and should be considered when designing regulatory frameworks, insurance coverage, and information architecture.

Hence, we hypothesize that there is a substantial amount of individuals that seek a resolution of uncertainty, i.e. seek a secured diagnostic result even though it does not lead to improved treatment, and therefore to a utility increase in the classical sense. This is a, if not the core hypothesis of our research, as it would be the first evidence for pure uncertainty resolution in healthcare.

H6: Some individuals seek information on their health, even though there is no direct utility benefit to it.

Further, in combination with the above-discussed mixed behavior regarding symptom levels, we hypothesize that just like increased symptoms do not increase the willingness to pay, increased symptoms do also not increase the preference for uncertainty resolution.

H7: With increasing symptoms, patients' preference to resolve uncertainty does not increase.

Social validation The research on the effect of social validation in health settings is limited but shows a consistent theme of its importance. Harmens et al. (2022) show the key role of being accepted by others for women with autism. Lillrank (2003) finds that social validation or social acceptance is one key reason, next to getting treatment and thereby pain relief in women with back pain. Further, in the case of EMS, a (factual) diagnosis could be psychologically highly relevant to many patients as it creates acceptance for themselves and their environment (ESHRE GDG, 2022). We hypothesize that social validation is the key driver to taking the test, in cases in which it gives no benefit in treatment.

H8: Social validation is the key driver for uncertainty resolution.

Chapter 3

Methods

In the following, we discuss the ethical concerns of our research, followed by the experimental design, thereby reasoning why we leverage a Discrete Choice Experiment (DCE) as a tool to answer our research question. We then detail our sample requirements and discuss the actual sample. Following, we walk through the experimental procedure that participants experience and discuss all recorded variables and their functions. Finally, we discuss how the collected data is analyzed.

Ethical Approval Due to the medical nature of this experiment and due to the core location being German, as the decision-making design is adapted to Germany's healthcare market, ethical concerns arose in the approval process. The choice of Germany as the core location of the research builds upon the combination of multiple factors that enable a realistic choice architecture to be displayed in the DCE. Firstly, patients in Germany are not required to pass their General Practitioner, who may act as a gatekeeper. Given that EMS is strongly underdiagnosed and many, especially non-specialized, practitioners lack knowledge (Nnoaham et al., 2011), combined with the choice architecture of the experiment only accounting for one appointment, it is key that patients can decide to directly consult a gynecologist or a doctor specialized in EMS. Secondly, doctors appointments not coming with a cost is key in ensuring that the decision of visiting the doctor does not depend on financial viability. Lastly, the availability and clear pricing situation of the diagnostic test in Germany are key to representing a realistic choice architecture.

Next, the medical nature of this experiment requires controlling for patients' actual experience of symptoms. To ensure doing this in an ethical manner questions from the well-established *EQ-5D* from the Dutch *EuroQol Research Foundation* were used, which have been applied for more than 30 years (Brooks & De Charro, 2003; Derrett et al., 2016; Gusi et al., 2010). More specifically, only the relevant categories of *usual activities*, *pain/discomfort*, and *anxiety/depression* were used, limiting the responses to the relevant symptoms. Hence, well-tested and established measures were used to minimize ethical concerns.

Finally, the experiment was assessed critically with the guidelines of the self-assessment questionnaire of the Ethics Commission of the HAW Hamburg and the *Research ethics principles and review procedures* by the German *Commission for Economic and Social Data*, which is backed by the German government (HAW Hamburg, 2022; RatSWD, 2017). This assessment ensures also meeting the ethical guidelines of Germany.

3.1 Experimental Design

As laid out in the previous section, our aim is to better understand patient preferences. Generally, there are three decision situations in which a patient may be. First, she might think she has EMS, e.g., due to a hereditary predisposition, without actually experiencing symptoms. In this case, she may proactively seek the diagnosis to possibly intervene in the development of EMS with early treatment. Second, she might experience symptoms but not be aware of EMS as a disease. In this case, she may seek information to understand the origin of the symptoms. Thereby, she could gain awareness of EMS in the information-seeking process, e.g. through the opinion of a general practitioner or a friend. Lastly, she may experience symptoms and already be aware of EMS being a possible origin.

Considering the limitation of our research to women above 18 years old, the first case is unlikely to be realistic, as EMS symptoms already develop in the first years of fertility (WHO, 2023). In the last two cases, women typically reach a stage in which they experience symptoms and have been made aware or already were aware of EMS as a possible cause. A well-educated general practitioner should refer a patient to a specialist and possibly inform her about the diagnostic test, following the latest standards for diagnosis as described in section 2.1. This is the stage of the decision situation that we portray. While it may be interesting to understand further decisions on patients with EMS risk or fear to have EMS, it exceeds the frame of research regarding Uncertainty Resolution and would dramatically increase the complexity.

Consequently, we want to understand if and how patients seek diagnostic information that enables treatment in optimal decision environments. This means patients are generally knowledgeable about EMS and can choose between the two mentioned diagnostic pathways that fulfill the latest standards, namely the diagnostic test or a specialist practitioner, or, alternatively, choose not to seek a diagnosis at all. As explained in detail in section 2.4 the probability of correct treatment still depends on many factors and may vary between specialists and within diagnostic efforts of the same specialist, i.e. Specialist-by-case interaction (Kahneman et al., 2021). Hence, trade-offs are between the test at different co-pay levels (the share that the insurance doesn't cover), specialist practitioners at different likelihoods of correct treatment, and not seeking any diagnosis. General practitioners are not an option as they should not be a part of the decision situation at this stage due to a lack of qualification and technological equipment for the suspicion diagnosis (ESHRE GDG, 2022).

To evaluate choice preferences in healthcare settings there is a variety of different tools. The most common tools can either be classified as revealed preference tools or stated preference tools. Revealed preferences tools observe actual behavior and derive preferences that are (indirectly) revealed through the choices. This is often preferred as there is no doubt that real preferences were captured. However, this is not always feasible. In healthcare markets goods are often not traded explicitly and there exists an agency relationship between the doctor and the patient suffering from asymmetric information and uncertainty in health state and outcomes. Hence, the revealed decisions may not actually portray the real preferences of an individual. This is where stated preference tools can be of higher value. They allow for the exact specification of the hypothetical choice setting, thereby enabling the clear identification of the effects of exogenously varied variables. Lastly, stated preference tools are rather cost-effective through comparatively simple recruitment (Ryan et al., 2008).

The contingent valuation method, which directly asks for participants' preferences, is one of two well-established methods of stated preference elicitation. However, directly asking individuals for their, e.g, willingness to pay can suffer from biases. Further, it is limited when leaving the realm of neoclassical welfare economics, i.e., when seemingly irrational behavior comes into the equation. Further, when exploring variables that individuals are not familiar with, e.g. prices of healthcare measures or exact accuracy of treatments, it is hard for them to explicitly name their preference (Ryan et al., 2008).

The second well-established method, the Discrete Choice Experiment (DCE), is better equipped to tackle such questions that are partially unfamiliar to individuals and include preferences driven by not only the direct economic benefit itself but also aspects such as anxiety and reassurance. Failure to measure such indirect preferences could lead to policy conclusions that do not fit the actual preferences of the population. In a DCE the participant can decide in a trade-off between two or more choice options in a choice set. Each choice is defined by specific attributes that can vary between the choice options. Typically a participant indicated their preference for multiple such choice sets, which then allows estimating preferences for the different attributes. This method is also less prone to biases and more accessible to people not familiar with the subject matter (Ryan et al., 2001; Ryan et al., 2008; Soekhai et al., 2019).

Lastly, the DCE benefits from its adaptability. E.g., one can repeat this study varying other attributes like the time until diagnosis, replicate the study with real patients, another illness, or unlabeled choice options to increase the understanding of patient behavior while remaining comparable between the studies. In this study, we opt for labeled alternatives for multiple reasons. Firstly, labeled alternatives enable the patient to easier understand the complex topic better and to make decisions that are closer to their decisions in the real world. E.g., with medication, the negative effect might be larger, as to most patients the names of medications mean nothing but may induce bias, while it does not increase understanding. Secondly, labeled alternatives enable using mixed-logit models which is key to our analysis, as will be discussed in section 3.5 and chapter 4. thirdly, if some of the alternatives without labels are still identifiable based on their attributes, while others are not, not using labels actually weakens the results (Ryan et al., 2008). E.g., different transport options are clearly distinguishable, traveling multiple hundred kilometers in a short time is obviously done by plane, and having many stops in a short trip is obviously public transport. The same applies here, as zero percent of treatment accuracy, no waiting time, and no diagnostic result clearly implies not seeking a diagnosis, while a free-of-cost option clearly implies a typical health insurance-covered diagnosis through a doctor, where no label would just lead to the participant making assumptions about the procedure, which in many cases could include the widely spread but outdated knowledge of a surgical intervention being needed for EMS diagnosis, as it used to be recommended in the treatment guidelines of 2014 (ESHRE GDG, 2022).

We opt for a "forced" choice experiment following Ryan et al. (2008), i.e. offering trade-offs in which the participant has to make a choice and cannot opt out of the given options. While e.g. forcing people to decide between two medications with side effects may lead to revealed preferences that do not portray realistic choice behavior, this seems very unlikely in the case of diagnostic choice, especially given that we offer the choice to avoid diagnosis in some trade-offs. Further, a forced choice enables a simpler survey design and reduces complexity for the participants.

3.2 Sample

Empirical basis To date, sample size calculations for DCEs are based on estimations and rules of thumb or are even disregarded entirely. a review of healthcare-related 2012 studies showed that merely 6% used parametric approaches while 13% used rules of thumb, while the rest simply did not report or referred to prior studies to justify their sample size. Further, they show that about a third of the reviewed studies use under 100 samples (de Bekker-Grob et al., 2015). One often-used rule of thumb used is by Orme (2019), who calculates a factor based on the intent to represent every main-effect level at least 500 times. However, he also notes that, for investigative work, 30 to 60 samples may suffice. Based on their empirical experience Lancsar and Louviere (2008) suggest that just 20 respondents are needed per questionnaire version if one aims to estimate a reliable model, however, extensive post hoc analysis may need a larger sample size. Pearmain and Kroes (1990) suggest that a total of 100 samples are needed to enable the modeling of preference data.

Parametric estimation To strengthen our decision we follow the established parametric approach by de Bekker-Grob et al. (2015) to estimate the critical effect size. While the sample size depends on the individual tested hypotheses, we will not specify the calculation for every hypothesis but detail the estimates for the general sample size (which regards topics such as willingness to pay and accuracy of treatment), as well as for the case of uncertainty resolution, as this is our main topics of interest. This approach relies on the following factors. First, it is designed for multinomial/conditional logit models of analysis. Second, we set the significance level α to 5% and the statistical power level $(1-\beta)$ to 80%, following de Bekker-Grob et al. (2015). Further, it depends on the DCE design (number of coefficients, alternatives, and choices) and design efficiency. The design chosen for our study is not necessarily efficient for all attributes but focuses on the main attributes of interest and hence is highly efficient for these attributes. Lastly, the estimation requires input regarding the initial belief about parameter values. We source these beliefs from an analysis of a test run of our questionnaire with 21 participants. Building on the parameters from this test run (*-0.001* for an additional euro, *0.02* for an additional percentage of accuracy, and *0.24* for resolution of uncertainty), the general model demands a sample size of 21, 7, and 37 samples for the attributes of price, accuracy, and uncertainty resolution respectively. Focusing exclusively on the trade-offs for uncertainty resolution, the parameters of *-0.001* for an additional euro and *0.47* for the resolution of uncertainty result in a sample size requirement of 61 and 39 samples respectively.

For a few reasons, we aim for a substantially higher sample size than the above calculated. For one, the uncertainty resolution parameter used to calculate the effect size is large but not significant. For another, a small but meaningful effect size may not be detected in a minimal sample. Together with the fact that sample size calculations of DCEs are always just estimates, we aim to double the needed effect size for the total sample, thereby ensuring that the individual groups of *Medium-Low* and *Medium-High* fulfill the requirements in themselves, and thus still guaranteeing meaningful results even if, e.g., many low-level participants would not see the need in treatment, perceiving the symptoms described as not significantly differentiating from what they perceive as normal for their period.

Final sample Participants were recruited through different social media channels including Instagram, Facebook, LinkedIn, Reddit, and Whatsapp. The survey recorded 393 answers, of which 18 were not female or under 18 years old. 75 individuals did not complete the survey, hence the final sample consists of exactly 300 participants. 253 of these are insured in Germany, of which 34 (9,6%) are privately health insured, which is close to the overall German population with a share of 10.7% privately insured, considering the young target group and that private insurance is only available with a certain income or employment status (BPB, 2022). 276 are under 45 years old and 237 are under 35 years old. This is relevant as most EMS patients are diagnosed by their late 20s (ESHRE GDG, 2022). The education was skewed towards highly educated people, with 84.6% having some form of university education, while only 31% of people in Germany have a university education and about 55% of high school graduates currently go to university (OECD, 2022; Statista, 2023). The income does not exactly represent the population but is not substantially skewed (IW, 2023). It is likely that students without income and graduates with over-average income balance the overall picture, to a certain extent. 14.4% of the sample reported that they have been diagnosed with EMS, which is above the average population, especially given that a substantial part of the estimated 6-10% of EMS patients has not been diagnosed to date (ESHRE GDG, 2022).

3.3 Procedure

In the following, the normal participation procedure is described in a concise manner. A full print of the survey can be found in appendix A.1. Participation was exclusively online, through Qualtrics (2023). There was no compensation or obligation to participate. The median participation time was 7 minutes and 50 seconds, with the vast majority of responses taking between four and a half minutes and 15 minutes.

Participants typically stumbled upon the survey on their social media feeds. Potentially, the survey link was shared with them directly through other survey participants. On social media variations of calls for participation in a study related to female health were made. Direct mentions of EMS were avoided to prevent bias in people's choice to participate depending on their knowledge of or attitude towards the disease. Again, in the introduction of the survey, the participants are not directly confronted with EMS but are informed that the questionnaire intends to understand the decision-making of patients better. They are asked to decide as they would in the real world. On the same page, participants are informed about their rights and are asked to actively acknowledge and consent to participation by choosing "yes". They are informed about the confidential and anonymous treatment of their data exclusively for research purposes as well as that their participation is voluntary and can be ended at any time without penalty. Finally, we displayed the contact details of the author.

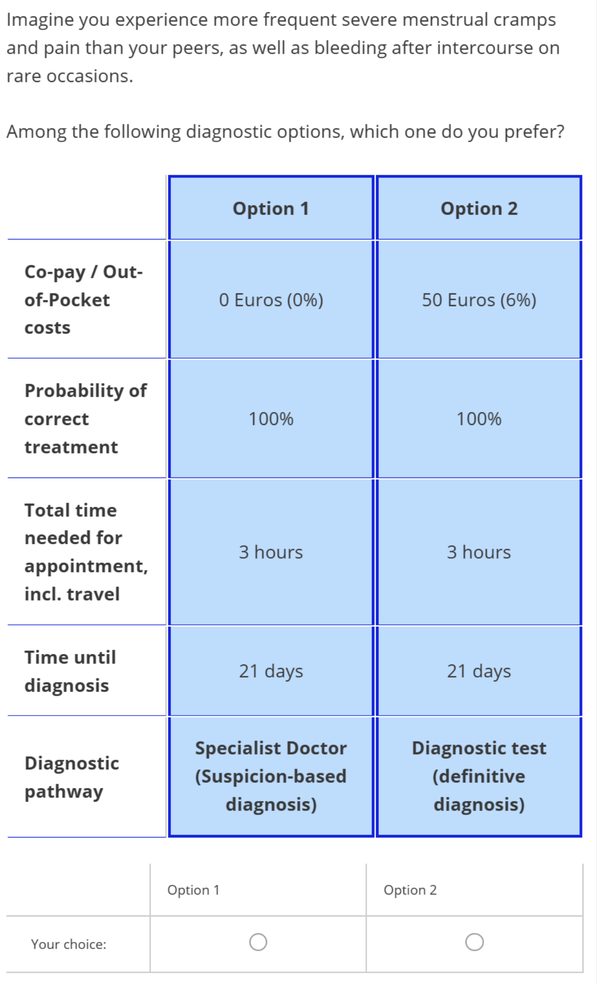
In a first step demographic information that qualifies for participation is inquired. This includes gender, age, and the type of health insurance in Germany or alternatively the country where the participant is insured. Participants of minor age or non-female participants are filtered out and directly sent to a personalized debriefing of the survey. As this filtering was key, answering the questions regarding gender and age were mandatory. In the second step, further demographics are collected, namely household income and educational attainment.

Next, the central part of the survey starts. Participants are informed about EMS, the sensible choices available for diagnosis and their up- and downsides, as well as typical symptoms, typical treatment, and social impact. Following they were presented with 10 choice sets of two alternatives each between which they must make a trade-off decision.

At this step, participants get randomly assigned into one of two groups. Both groups will receive the same 10 choice sets, however, one group will be provided with the *medium-low* and one with the *medium-high* description of symptoms that may indicate EMS, as elaborated in more detail in section section 2.1. The choice sets demand making trade-offs between two choices with five attributes, namely co-pay (out-of-pocket cost), the probability of correct treatment, the total time needed for the appointment, the time delay until having results of the diagnosis, and the diagnostic pathway. Participants decide which of the alternatives they prefer based on the given attributes. To better illustrate the setting, an example trade-off in the exact design participants saw is displayed in fig. 3.1. This example is for an individual that has been allocated to the *medium-low* symptom group and hence includes the respective description of symptoms. This description gets displayed in every trade-off.

All trade-offs the participants go through are summarized in table 3.1. The order of the trade-offs and the display sides of the choice options in the trade-off are randomized to minimize any kind of structural learning effects and biases, following Ryan et al. (2008). Hence, the displayed order is simply chosen for a better overview. The co-pay is set in euros and relative values, the probability of correct treatment in percent, the time needed in hours, and the time until getting diagnostic results in days. Lastly, we name the diagnostic pathway using the real description including the description of the diagnosis, hence "No diagnosis", "specialist Diagnosis", or "Diagnostic test". As the time needed and the time until receiving diagnostic results does not vary, except for the cases of no diagnosis (where they logically do not apply) we do not display them for better readability. Hence the annotation is: the diagnostic pathway with description in the top line, and the co-pay in Euros (relative values in parentheses), and the probability of correct treatment in percent in the second line.

As visible in the table 3.1 we prioritize certain aspects and combinations of attributes and do not vary all attributes in all choice options. This has two reasons. For one, creating choice options that would paint the option worse than in any real-world situation would be unethical as discussed in chapter 3 (HAW Hamburg, 2022; RatSWD, 2017). E.g., a choice option of a diagnostic test with 80% treatment accuracy could create the picture that this may be a factually possible outcome and hence drive people away from this option when actually considering diagnostic options for EMS in real life. This type of choice set would be misleading and could lead to worse health outcomes and is therefore not possible following ethical standards. For another, we designed the trade-offs focusing on the topics of interest in our hypotheses. E.g., evaluating the effect of different waiting times was not part of the research scope. Further,



Imagine you experience more frequent severe menstrual cramps and pain than your peers, as well as bleeding after intercourse on rare occasions.

Among the following diagnostic options, which one do you prefer?

	Option 1	Option 2
Co-pay / Out-of-Pocket costs	0 Euros (0%)	50 Euros (6%)
Probability of correct treatment	100%	100%
Total time needed for appointment, incl. travel	3 hours	3 hours
Time until diagnosis	21 days	21 days
Diagnostic pathway	Specialist Doctor (Suspicion-based diagnosis)	Diagnostic test (definitive diagnosis)

	Option 1	Option 2
Your choice:	○	○

Figure 3.1: Example choice set for trade-off questions of DCE

testing more combinations of prices of the test at the given accuracy levels of the specialist, like 200€ cost for the diagnostic test at 95% specialist accuracy, would not have added substantial value, especially considering the hypothetical nature of the decision situation, as the exact price point is unlikely to hold in real situations.

Finally, participants are asked to answer questions about their EMS-symptom-related health status, their attitude about doctors visits, and their motivations to take the diagnostic test given 100% accuracy of treatment.

Debriefing If participants gave their consent to participate there were two possible paths for the debriefing. If they did not fit the minimum criteria to participate, namely being above 18 years old and female, they were filtered out and shown a message that thanked them for their participation and explained why they cannot continue (see A.2). If they did match the criteria participants were equally thanked for their participation and given reliable sources to learn about EMS or find highly qualified doctors. In both cases, contact details were displayed in case they may have questions or concerns.

Table 3.1: Trade-offs in choice tasks

Alternative 1	Alternative 2
No diagnosis 0€, 0%	Specialist Doctor (Suspicion-based diagn.) 0€, 80%
No diagnosis 0€, 0%	Specialist Doctor (Suspicion-based diagn.) 0€, 100%
No diagnosis 0€, 0%	Diagnostic test (definitive diagnosis) 200€ (25%), 100%
No diagnosis 0€, 0%	Diagnostic test (definitive diagnosis) 800€ (100%), 100%
Specialist Doctor (Suspicion-based diagn.) 0€, 80%	Diagnostic test (definitive diagnosis) 200€ (25%), 100%
Specialist Doctor (Suspicion-based diagn.) 0€, 80%	Diagnostic test (definitive diagnosis) 800€ (100%), 100%
Specialist Doctor (Suspicion-based diagn.) 0€, 95%	Diagnostic test (definitive diagnosis) 50€ (6%), 100%
Specialist Doctor (Suspicion-based diagn.) 0€, 99%	Diagnostic test (definitive diagnosis) 50€ (6%), 100%
Specialist Doctor (Suspicion-based diagn.) 0€, 100%	Diagnostic test (definitive diagnosis) 50€ (6%), 100%
Specialist Doctor (Suspicion-based diagn.) 0€, 100%	Diagnostic test (definitive diagnosis) 800€ (100%), 100%

3.4 Variables

In this section, we discuss all variables and their function. *Selected Choice* is our dependent variable used in the regressions to predict choices. Further, there are different groups of independent variables. For one, there are demographic variables that describe and enable us to account for the individual subject characteristics. Second, there are the exogenously varied variables of the discrete choice experiment itself, enabling the core part of our analysis. Third, the variables regarding medical self-reporting and attitude let us understand the effect of medical state and self-reported medical behavior. Lastly, we inquire about the motivations, hence the reasons why individuals take the diagnostic test in situations that may display pure UR.

Selected Choice For our model we create the dependant variable *Selected Choice*. In the Qualtrics data set, there is only a data point for each chosen option. To enable a choice data regression model, we create a data point for every choice option that was not chosen. The dummy variable *Selected Choice* is equal to 1 for the chosen option and 0 for the not chosen option. This enables predicting the choice, depending on the following independent variables.

3.4.1 Demographics

Age The variable age has multiple important functions. First, it suites as a filter for minor subjects as discussed in section 3.2. Second, EMS is a disease that typically onsets with fertility and worsens, to a certain extent, over time, if it is not treated (ESHRE GDG, 2022). However with the onset of Menopause symptoms normally decrease or disappear altogether (ESHRE GDG, 2022) which is typically around 45 (Ceylan & Özerdoğan, 2015). Given the onset with fertility and a typical delay of diagnosis of 8-12 years, most EMS patients with significant quality-of-life impact get diagnosed before 35 or even 25. Hence, these groups of participants are especially relevant.

Gender While this variable is collected, all participants that complete the survey identify themselves as female. All others get filtered out as specified in section 3.2.

Insurance We collected the insurance type in Germany, as this is the healthcare market design that the experiment reflects. Namely, participants indicated statutory (public) health insurance (SHI) or private health insurance (PHI). If the participant was not insured in Germany they could indicate their country of insurance.

Education Educational attainment was collected in a detailed variety from "some primary school" to "graduate or professional degree", including the options to state some education of a certain level, without obtaining a degree to date, thereby ensuring a better understanding of the educational level of many young adult participants.

Household Income Income is collected as a gross income on the household level per year, in standardized categories of smaller than 25000€, 25000€ - 49 999€, 50 000€ - 99 999€, 100 000€ - 199 999€, and over 200 000€. These categories are "Qualtrics Certified: standardized Region Specific Demographics" created and tested by Qualtrics experts (Qualtrics, 2023).

3.4.2 Discrete Choice Experiment

The following attributes, as well as the randomly assigned symptom levels, are exogenously controlled variables, i.e., they are predetermined and are consciously chosen to drive the participants' decisions.

Randomization into symptom-level groups Participants are randomly allocated into two levels, *medium-low* and *medium-high* symptoms, as described in the following. Participants are asked to imagine actually suffering from these symptoms and to make their decisions based on this hypothetical situation.

To identify a potential effect of the severity of symptoms, two levels were created. Hereby it is key that the symptoms describe realistic choices that are relevant in this choice setting. For one, they may not belong to the cases of very high levels of severity, as, e.g., having seven symptoms already increases the odds ratio of being diagnosed with EMS to 84.7 (95% CI 58.8 to 121.8)(Ballard et al., 2008), and would often entail surgery as part of the treatment, making an uncertain suspicion diagnosis obsolete. On the other hand, asymptomatic cases do not necessarily demand treatment (ESHRE GDG, 2022).

While the assessment of the latest scientific knowledge and therefore treatment recommendations differ, We base all descriptions and knowledge shared with study participants on *The Endometriosis Guideline of European Society of Human Reproduction and Embryology* (ESHRE

GDG, 2022) to ensure a coherent picture for participants. Building on these European treatment guidelines the following two descriptions were developed to display a case of significant but rather ambiguous symptoms from here on referred to as *Medium-Low (L)* and a case of severe, more clear symptoms with a defined impact on quality-of-life referred to as *Medium-High (H)*. The defined symptoms are based on the diagnostic guidelines of the ESHRE and were confirmed with experts and patient interviews.

Medium-Low *Imagine you experience more frequent severe menstrual cramps and pain than your peers, as well as bleeding after intercourse on rare occasions.*

Medium-High *Imagine you suffer from chronic pelvic pain that worsens before the onset of menstruation. Further, you experience stronger and more frequent severe menstrual cramps, pain, and bleeding than your peers as well as occasional bleeding after and pain during intercourse. These symptoms impact your overall quality of life and occasionally your ability to work normally.*

Co-pay This key attribute describes the share of the cost a patient must cover from her own pocket, i.e., what their insurance does not cover. The cost is stated in absolute and relative value. This attribute only varies for the diagnostic test, as a practitioner visit is not connected to cost in the German healthcare market. Of course, choosing "no diagnosis" does not come at a cost either.

Probability of correct treatment This key attribute describes the likelihood of receiving the correct treatment after a diagnosis of the indicated type (see diagnostic pathway). The likelihood does not describe if the diagnosis was actually correct. As described in more detail in section 2.1, the treatment of possibly EMS-related symptoms can be rather generic, therefore potentially being accurate independent of a correct diagnosis. This attribute does not vary for no diagnosis as it is assumed that no treatment will be done, and it does not vary for the diagnostic test as the accuracy of the diagnostic test is not subject to skill or any other fluctuating factor. It varies for the specialist practitioner dependent on the factors such as the skill level, as discussed in section 2.2.

Time needed This attribute describes the total time needed for the appointment, including travel time to and from the appointment. It is hence the time that an individual has to actively invest out of their day to enable choosing this option. This attribute stays fixed and only varies if the participant decides against a diagnosis, in which case of course there is no time investment.

Time until diagnosis This attribute describes the time delay between deciding on a diagnostic pathway and the moment of getting the results. For the Specialist doctor, this includes the waiting time from making an appointment to receiving the result at the appointment. For the diagnostic test, this includes the time waiting for the test appointment and the ensuing time waiting for laboratory results. Again, this attribute stays unchanged and only varies if the participant decides against a diagnosis, in which case there is no diagnosis and hence no waiting time for the results.

Diagnostic pathway This attribute is at the same time the description of the choice option and the key differentiator for the factual certainty of knowledge. It includes the options "no diagnosis", "specialist doctor (Suspicion-based diagnosis)", and "diagnostic test (definitive diagnosis)".

3.4.3 Medical self-reporting and attitude

The following variables record the actual health status and health attitudes of participants. The self-reported health status enables us to account for the influence already having been diagnosed or actually suffering from the symptoms described has on the stated preferences of individuals. Further, by inquiring about the individuals' attitudes towards healthcare-seeking and medical professionals, we can understand better what might drive patients away from a diagnosis. Note that a variable referred to as "symptoms" was created for the analysis that includes self-assessment of pain and discomfort caused by symptoms, anxiety, and the extent to which these symptoms impact usual life activities. The variable is the sum of the three indicated self-assessments but is set to 0 if not at least pain and discomfort pr impact on usual life activities is described as moderate, to exclude people that only experience slight or no pain, discomfort, or impact on their everyday life, as well as exclusively anxiety, as this is a very broad indication.

Endometriosis This is a binary variable that describes if the participant self-reports having been diagnosed with EMS by a medical professional. This functions as a control, to see whether already diagnosed patients may be biased in their decision-making.

Anxiety and depression This attribute controls for actual symptoms that may be related to underlying, undiagnosed, EMS. Anxiety and depression are typical symptoms that may come with EMS (ESHRE GDG, 2022). Hence, it is key to control for them. The question is based on the EQ-5D-5L scoring from EuroQol, which is a well-approved quality-of-life measure (Derrett et al., 2016). While anxiety itself is not an indicator, a combination with the following two variables makes it a significant predictor of EMS. Further, as discussed in section 2.4 anxiety can drive people to avoid information.

Usual activity This attribute controls for the impact on the ability to perform daily activities of a list of typical symptoms that may be related to underlying, undiagnosed, EMS. Again, the question is based on the EQ-5D-5L scoring from EuroQol (Derrett et al., 2016), which is a well-established quality-of-life measure. We limit the list to a selection of the most typical EMS related symptoms. This is an act of balance to keep the questionnaire simple, short, and understandable for the participants while enabling to control for the impact of actually experiencing the symptoms described, in contrast to those participants that do not experience symptoms that may be related to EMS.

Pain and discomfort As a third predictor of EMS, this describes to which degree participants experience pain or discomfort originating from symptoms that are typical for EMS. Like the variable *usual activity*, this is based on EQ-5D-5L scoring and enables to account for actual experience of possibly EMS-related symptoms.

Attitude: *Low perceived need* As discussed in detail in section 2.4 some people tend to avoid seeking medical professionals as they have a low perceived need, thinking a doctor's visit as unnecessary. To account for this behavioral preference, we ask for people's attitudes about their health, i.e., if they tend to seek medical care quickly or only when really necessary.

Attitude: *Discomfort* In one study Taber et al. (2015) find that 27% of participants avoid seeking medical care due to feeling uncomfortable. We measure this behavioral factor with a subjective self-assessment, from being very comfortable to avoiding doctor visits due to discomfort and anxiety.

Attitude: *Uncertainty resolution self-report* This variable acts as a control for a subjective, self-assessed preference for uncertainty resolution. If people choose the diagnostic test in cases of equal outcomes but do not score highly on this measure, this would indicate that other reasons, like discomfort related to imaging technologies, may drive their choice, or they are not aware of their preference to resolve uncertainty.

3.4.4 Reasons to take the diagnostic test

These variables inquire for subjective reasons that may motivate participants to take the diagnostic test, given no treatment benefit in comparison to the specialist practitioner. The list given supplies the two main reasons elicited in preceding qualitative interviews with women, EMS patients, and experts, as well as research (ESHRE GDG, 2022; Lillrank, 2003). Participants can score these reasons on a scale from irrelevant to highly relevant, hence having free choice. E.g., they can score all of the reasons low if they do not see a need in doing the diagnostic test at all. Factual knowledge and social acceptance are potential drivers for uncertainty resolution but patients may name reasons that indicate other preferences.

Factual knowledge Participants score how important having a factual confirmation of the diagnosis is to them in assessing if they should take the test, even if there is no treatment benefit.

Social Acceptance Participants score how important creating social acceptance for themselves or in their environment is to them in assessing if they should take the test, even if there is no benefit.

Other This variable, including a text entry field, allows participants to score any other reason(s) that would influence them to take the diagnostic test and share the specific reasons if they wish.

3.5 Analysis

To analyze the data we draw back to well-established tools for DCEs. After data cleaning as described in section 3.2 we first start with creating a dummy variable for the selected choice as explained in section 3.4. Then we gain an overview with a short descriptive analysis. thereafter, we start our general analysis using McFadden's multinomial logistic regression model (MNL) (McFadden et al., 1973).

Using MNL is a common if not the most common practice in DCEs, together with conditional logit models (De Bekker-Grob et al., 2015). While researchers disagree if MNL and conditional logit models are indifferent, they are inherently the same in their foundation and the terms are often used interchangeably in stated choice experiments (Ryan et al., 2008). The four assumptions of MNL are as follows. (1) the independence of irrelevant alternatives (IIA), (2) homoscedasticity, (3) the respondents must have homogeneous preferences between each other, and (4) there may not be a correlation within responses. Rightfully one might question if these assumptions hold. While the MNL is computationally light, these restrictions make it less optimal for our analysis.

One can use the mixed logit model to allow for correlation in unobserved factors over time, unrestricted substitution patterns, and random taste variation (McFadden & Train, 2000; Train, 2001). Thus, we perform a mixed logit regression fit to choice data for repeated choices (i.e. a panel data structure) (Rabe-Hesketh & Skrondal, 2022). This accounts for the choice task (therefore the correlation between responses) and the choice alternative and enables a complete IIA relaxation (De Bekker-Grob, 2009). Unlike the MNL, this panel-data mixed logit model (PMLM) models the probability of choosing a specific alternative in a specific choice task and therefore is inherently superior to the MNL for our analysis as it also allows for a more specific interpretation of choice behavior dependent on labeled alternatives (the labels are "Specialist doctor", "diagnostic test", and "no diagnosis").

We hence build on the PMLM in our results after ensuring the estimations for both the MNL and PMLM are similar, indicating model validity. Finally, we will check for the robustness of the MLM based on sample composition and other factors.

While one might argue that probit models are also well fit to analyze DCE data, they do not always match the experimental design choices (Rose & Bliemer, 2009). For a probit model the attributes must be set up to be as equally distributed as possible. This is not the case in the given experiment, as not an equal distribution of attributes but certain key attributes were prioritized, and not all types of combinations were used (e.g. a test with low treatment accuracy does not occur as it is not a realistic option). Fitting the experiment optimally for a probit model would not add value as choice preferences that do not exist in the real world would be elicited. Contrarily this could create ethical issues. Further, there is substantially less empirical backing for probit models compared to the commonly used MNL models in healthcare settings and DCEs in general, as well as for specifics such as DCE design and sample size calculation, which would weaken the results.

For comparison tests of means, we use t-tests, as we assume a random sample with homogenous variances and an approximately normal distribution, given the large, random sample. Similarly, we use multivariate tests of means when comparing multiple means.

Chapter 4

Results

This chapter focuses on the DCE but includes additional questions on health state, motivation, and reasoning. we lay out the results of the questionnaire results and connect these to the hypotheses stated in chapter 2. We first give a descriptive overview of the results and then look into the model results respectively.

Additionally to the main sample, we created a control group that exclusively contains women under 35 that live in Germany and that answered all questions, thereby ensuring the allocation of "prefer not to say" answers does not impact the results. We will run our analysis with the larger group of 300 participants, and control for robustness with this control group of 158 participants. With this control group, we represent the actual sample group of interest, as we focus on the design of the German healthcare market, and patients with EMS that significantly impact their quality of life are typically diagnosed by 35. In the table 4.1 one can find an overview of the participant population, for the entire population and the control group respectively, in absolute and relative terms. For the Regression models detailed later, we group education into two groups of higher and lower educational attainment, wherein everyone that has finalized or is currently pursuing a university degree is classified as having higher educational attainment. This is done as many participants are still of university-typical age and thus have not finished their degrees. Hence grouping them into a lower education group than people further in their careers does not make sense, rather summarizing people having attained or striving for higher attainment better represents the true picture.

4.1 General model

Looking at the estimated models, in the first step we want to observe the general results and ensure the validity of the PMLM estimation method. In the second step, we want to choose the best model based on goodness of fit measures. As discussed in section 3.5 it is common practice to first estimate an MNL as the base and then compare the estimates to those of a (panel-)MLM, wherein similar coefficient signs and significance levels prove the validity of the model. Thereafter, one can assess and improve the best model based on goodness of fit criteria. We focus on, and thus report, the log-likelihood, the Akaike information criterion (AIC), and the Bayesian information criterion (BIC), as is common practice for MLM models (Müller et al., 2013).

Table 4.1: Summary Statistics

Sample	Entire sample		Control sample	
Variable	Frequency	Percentage	Frequency	Percentage
Household Income (€/year)				
Less than 25,000	100	33.33	73	46.20
25,000 - 49,999	73	24.33	44	27.85
50,000 - 99,999	71	23.67	32	20.25
100,000 - 199,999	28	9.33	9	5.70
More than 200,000	2	0.67		
Not specified	26	8.67		
Age				
18-24 years old	98	32.67	54	34.18
25-34 years old	139	46.33	104	65.82
35-44 years old	39	13.00		
45-54 years old	11	3.67		
55-64 years old	10	3.33		
65+ years old	3	1.00		
Education				
Lower educational attainment	46	15.33	22	13.92
Higher educational attainment	254	84.67	136	86.08
Insurance				
PHI	34	11.33	21	13.29
SHI	219	73.00	137	86.71
Non-german	47	15.67		
User language				
German	243	81.00	138	87.34
English	57	19.00	20	12.66

Following step one, we estimate a base model with only the exogenously varied variables, and a complete model with all available variables, respectively for the MNL and PMLM estimation. Models 1 and 2 in table 4.2 hence show the MNL which is fit for panel data (the multiple choice tasks) per individual on the selected choice variable. Model 1 only contains the exogenously varied variables, hence the co-pay, accuracy of treatment, and symptom level. Model 2 adds all variables, including explanatory (e.g. healthcare attitude) and control variables (e.g. age).

The base model, model 1, predicts the dependant variable *Selected Choice* that was created for the regression model as described in section 3.4, using an MNL regression as laid out in section 3.5. The main explanatory variables are the exogenously varied attributes of interest. As we only vary the co-pay, the probability of correct treatment, the diagnostic pathway, and the symptom level between specialists and diagnostic tests, we omit the other attributes supplied in the DCE. While they do change with the choice option of *no treatment*, they are not separable from the diagnostic pathway attribute, due to the nature of the choice never having waiting time or time needed for the appointment. Model 2 reports the extended model with all independent variables. It is, as expected less efficient as it cannot account for the alternatives of the trade-off.

Model 3 and 4 in table 4.3 represent the same variable composition as model 1 and 2 respectively but are regressions that fit the PMLM to choice data. As the models account for the

Table 4.2: Results of MNL Regressions

Predictor	Model 1	Model 2
No diagnosis	(base alternative)	(base alternative)
Co-pay	-0.0020***	-0.0020***
	(-0.0022, -0.0017)	(-0.0022, -0.0017)
Treatment accuracy	0.0296***	0.0296***
	(0.0270, 0.0321)	(0.0271, 0.0321)
Symptom level	-.0203	-.0200
	(-0.2090, 0.1685)	(-0.2122, 0.1721)
Age		-0.0096
		(-0.1282, 0.1090)
Private insurance		-0.0513
		(-0.4710, 0.3685)
Public insurance		-0.0073
		(-0.3209, 0.3062)
Higher education		-0.0079
		(-0.2852, 0.2694)
Household Income		0.0137
		(-0.0646, 0.0920)
EMS		-0.0205
		(-0.3153, 0.2744)
Anxiety		0.0039
		(-0.0934, 0.1011)
Actual symptoms		0.0075
		(-0.0229, 0.0379)
Low perceived need		-0.0103
		(-0.1068, 0.0863)
Discomfort		-0.0072
		(-0.0946, 0.0801)
Self reported UR		-0.0150
		(-0.0828, 0.0528)
Language		-0.0159
		(-0.3118, 0.2800)
Constant	-2.0387***	-1.9950***
	(-2.2685, -1.8089)	(-2.6770, -1.3130)
Log likelihood	-3468.649	-3468.184
AIC	6947.30	6970.368
BIC	6980.80	7084.259

Note. 95% confidence intervals are recorded in parentheses. Significance levels: * = .05, ** = .01, and *** = .001.

correlation of choices across alternatives (i.e. it allows for random coefficients), the IIA assumption of the MNL model is relaxed. Thus one also typically expects a better fit of the model. While the MNL, just like a usual logistic regression model, creates one single estimate for each variable's explanatory effect of choosing versus not choosing a trade-off alternative independent of the chosen alternative, the MLM creates individual estimators for the coefficients for which preferences vary randomly between the subjects (co-pay and treatment accuracy) and alternative specific estimates for the choice of the individual alternatives "no diagnosis" (here the base alternative), the "specialist diagnosis", and the "diagnostic test". Hence, the table is structured differently with an estimate per unit individual variable for each outcome. This means there is, e.g., an estimate of self-reported symptoms predicting the choice of the diagnostic test and a separate estimate of self-reported symptoms predicting the choice of a specialist practitioner, respectively compared to the chosen base outcome. This allows for more meaningful estimates and is a benefit of using labeled alternatives compared to unlabeled alternatives such as diagnostic choices "A" and "B" which do not allow for alternative specific covariates and thus alternative specific predictive margins.

The standard deviation reported under the header "Normal" displays how the coefficients vary in the population assuming a normal distribution and the correlation of co-pay and accuracy indicates to what extent these variations across the population are correlated with each other. Simply said a positive value would mean someone more sensitive to co-pay is also more sensitive to treatment accuracy. Accounting for this variation in individual preferences regarding co-pay and correlation improves the model, another benefit of the MLM.

Comparing the coefficients of models 1 and 3, and 2 and 4 respectively in tables 4.2 and 4.3 we can see that for all estimates the choice data fitted panel MLM is a valid, and thus well-suited, estimator as the results of the exogenously varied attributes in the trade-off design have the same signs and similar significance between the MNL and PMLM models. Comparing the goodness of fit criterion we can conclude that the PMLM models have a far superior fit and in themselves partially improve when adding variables, even if they are not all significant: The superior fit of PMLM is logical given they account for the choice task and individual preferences, hence correlation across alternatives. This also explains why unit individual variables like self-reported preferences for UR are not meaningful in the MNL. In model 4 one can note that language does not have a significant effect at 5% with a p-value of 0.74 for the choice of the diagnostic test and 0.71 for the choice of a specialist. No significant effect of the chosen language indicates that the translation was done well and doesn't bias the answers.

In the estimation process, we added variable groups (demographics, medical self-reporting, and attitudes) step-wise. Hereby we noticed that the majority of demographic variables did not improve the model fit. This is possibly due to the hypothetical setting, or due to a rather homogeneous sample group consisting of many highly educated individuals. Hence, in the last step, we optimized the model for goodness of fit, resulting in model 5. It includes all exogenously varied variables and all health variables, as these are predictors in our analysis. The only demographic variable included is household income, all other demographic variables are excluded as they do not (sufficiently) improve the model in this hypothetical scenario with the given, partly homogenous, sample. This model is the best fit for the data set and hence the model on which we will base all further references. Hence all models mentioned while answering hypotheses will be of this preferred model composition using a PMLM estimator, except if clearly stated otherwise. Looking at the results we that there is a positive effect of an additional percentage of treatment accuracy and a negative effect of an additional euro of co-pay cost on the selected diagnostic choice, ceteris paribus. Both effects are significant at 1%.

Table 4.3: Results of Mixed Logit Choice Models

Predictor	Model 3	Model 4	Model 5
No diagnosis	(base alternative)	(base alternative)	(base alternative)
Diagnosis			
Co-pay	-0.0079***	-0.0068***	-0.0070***
	(-0.0098, -0.0060)	(-0.0083, -0.0053)	(-0.0085, -0.0055)
Treatment accuracy	0.0671***	0.0530**	0.0546***
	(0.0450, 0.0891)	(0.0333, 0.0727)	(0.0346, 0.0745)
Normal			
SD Co-pay	0.0086	0.0063	0.0065
	(0.0068, 0.0108)	(0.0049, 0.0083)	(0.0050, 0.0084)
Corr. Co-pay x tr. acc.	0.1419*	0.3338**	0.3390***
	(-0.0016, 0.2796)	(0.0505, 0.5673)	(0.0957, 0.5441)
SD Treatment acc.	0.0393	0.0268	0.0285
	(0.0291, 0.0531)	(0.0187, 0.0383)	(0.0201, 0.0404)
Diagnostic test			
Symptom level	0.3368	-0.0898	-0.0995
	(-1.0477, 1.7212)	(-1.2556, 1.0760)	(-1.2738, 1.0749)
Age		0.4841	
		(-0.2682, 1.2364)	
Private insurance		1.9393	
		(-1.0584, 4.9370)	
Public insurance		0.3923	
		(-1.6808, 2.4653)	
Higher education		-0.3425	
		(-1.9469, 1.2619)	
Household Income		-0.3757	-0.4331
		(-0.8636, 0.1121)	(-0.8953, 0.0292)
EMS		1.1895	1.3080
		(-0.6332, 3.0122)	(-0.5455, 3.1615)
Anxiety		-0.1165	-0.1103
		(-0.6661, 0.4331)	(-0.6719, 0.4513)
Actual symptoms		-0.2677***	-0.2878***
		(-0.4571, -0.0783)	(-0.4769, -0.0987)
Low perceived need		0.3476	0.2717
		(-0.2639, 0.9592)	(-0.3449, 0.8882)
Discomfort		0.1021	0.0690
		(-0.4245, 0.6286)	(-0.4484, 0.5865)
Self reported UR		0.7790***	0.7866***
		(0.3057, 1.2523)	(0.3043, 1.2689)

Note. 95% confidence intervals are recorded in parentheses. Significance levels: * = .05, ** = .01, *** = .001.

Table 4.3 – *Continued from previous page*

Predictor	Model 3	Model 4	Model 5
Language		0.7421	
		(-1.2397, 2.7238)	
Constant	2.1343*	0.3809	2.1098
	(0.4032, 3.8655)	(-4.0215, 4.7832)	(-1.2306, 5.4501)
Specialist Doctor			
Symptom level	0.4149	0.0801	0.0648
	(-0.9035, 1.7332)	(-1.0209, 1.1810)	(-1.0435, 1.1732)
Age		0.4081	
		(-0.3086, 1.1249)	
Private insurance		1.6876	
		(-1.1751, 4.5502)	
Public insurance		0.3689	
		(-1.6091, 2.3470)	
Higher education		-0.4290	
		(-1.9446, 1.0865)	
Household Income		-0.3879	-0.4338
		(-0.8484, 0.0727)	(-0.8707, 0.0031)
EMS		1.4592	1.5822
		(-0.2815, 3.1999)	(-0.1896, 3.3540)
Anxiety		0.0481	0.0500
		(-0.4694, 0.5656)	(-0.4782, 0.5781)
Actual symptoms		-0.3371***	-0.3551***
		(-0.5174, -0.1567)	(-0.5356, -0.1746)
Low perceived need		0.4076	0.3419
		(-0.1738, 0.9890)	(-0.2436, 0.9273)
Discomfort		0.0843	0.0604
		(-0.4138, 0.5823)	(-0.4284, 0.5492)
Self reported UR		0.3071	0.3118
		(-0.1437, 0.7578)	(-0.1469, 0.7705)
Language		0.7901	
		(-1.1019, 2.6821)	
Constant	1.2193	0.6659	2.0595
	(-0.4359, 2.8745)	(-3.5061, 4.8378)	(-1.1063, 5.2252)
Log-likelihood	-1274.605	-1215.225	-1217.85
AIC	2567.21	2496.449	2481.7
BIC	2621.267	2694.659	2619.847

Note. 95% confidence intervals are recorded in parentheses. Significance levels: * = .05, ** = .01, *** = .001.

4.2 Willingness to pay for accuracy of treatment

With hypothesis one, we intend to test if most subjects choose the cost-minimizing, benefit-maximizing choice, hence behaving according to the predictions of neoclassical economic theory. As laid out in section 2.2, rational decision-makers should only take into account direct economic or personal value. Hence, if the health outcome is not improved, individuals should prefer the option with the lowest cost or be indifferent between equally costly options. However, taking a descriptive look at the stated preferences from the relevant choice tasks already indicates a different picture. While at a price point of 800€ for the diagnostic test and no difference in treatment accuracy, 243 out of 300 individuals (81%) chose the specialist practitioner, and therefore the optimal choice according to traditional theories, this also means that 57 (19%) did not. Looking at the same trade-off with a reduced price point of 50€, only 116 out of 300 individuals (39.67%) chose the specialist practitioner. Hence, 184 individuals, or 61.33%, did not act according to traditional economic theory. Verifying with a two-sided t-test, that this mean is indeed statistically different and larger than 50%, we can reject the null hypothesis (that it is equal to 50%), significant at a $p<0.001$ level, with $t(299)=-13.66$. This means that, at a price point of 50€ in the hypothetical DCE setting, a majority does not make the choice neoclassical models would predict.

Consequently, we reject our first hypothesis. Our experiment cannot confirm that a majority acts according to traditional economic models, thereby only optimizing their utility for direct benefits. More on why this might be the case is revealed in our analysis of hypotheses 7 to 9.

However, this leads directly to the second hypothesis. We expect the diagnostic test to be chosen more often with decreasing price (co-pay). Comparing the stated choices (i.e. the answers in specific trade-offs) in the DCE descriptively, 172 individuals or 57.3% choose the diagnostic test with 100% accuracy of treatment when given the alternative of treatment through a specialist at 80% accuracy. This share goes down to 26.7% (83 individuals) at a price of 800€ for the diagnostic test at 100% treatment accuracy keeping the specialist alternative constant. Looking at the coefficient in the preferred model (5), there is a significant negative effect per Euro cost of co-pay, on the selected choice, at $p<0.001$. This indicates that there is indeed a significant effect of price (in the form of co-pay) on the diagnostic choice made by individuals but it does not clarify the magnitude or the impact on the diagnostic pathway selected.

It is important to understand that the coefficient of PMLM models cannot be easily interpreted in its magnitude. Hence, to get a better understanding we analyze the predicted margins of the preferred model (5) at different price levels while keeping all other aspects fixed at mean levels. The predictive margins of PMLM models can be interpreted as the probability of an alternative being chosen under the specified condition, wherein all other variables are set to the mean observed values. This estimation includes all evaluated diagnostic choices, hence not only two as in the DCE choice tasks. Hence, in this case, one can interpret the coefficients as the percentage of individuals that choose the diagnostic test, the specialist diagnosis, and no diagnosis, respectively at a specified co-pay of 50€ for the diagnostic test and mean observed values for all other variables (this means every individual is seen as an average individual).

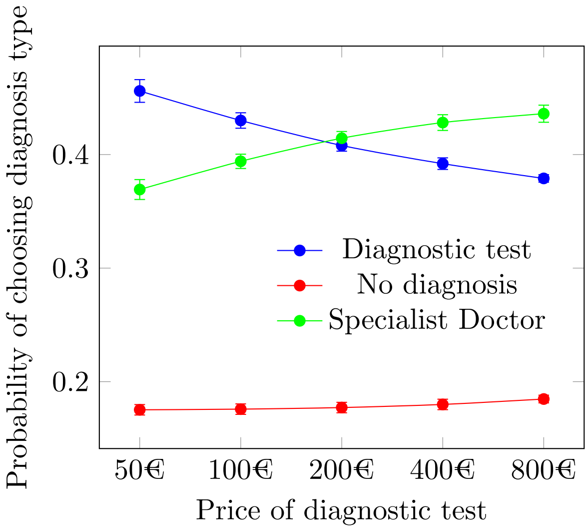


Figure 4.1: Predictive margins of diagnostic choice dependant on cost of the DT

The predicted margins for different price levels of the diagnostic test, while keeping all other variables fixed at mean levels, are displayed in fig. 4.1 with the corresponding 95% intervals. We can see that, while there is a small reduction, there is no substantial change in the share of individuals choosing no diagnosis with decreasing prices. However, there is a strong change in the share of individuals that chooses the diagnostic test which increases from 37.9% (CI: 0.373; 0.385) at a price-point of 800€ to 45.6% (CI: 0.446; 0.465) at price-point of 50€, and a contrasting decrease of individuals choosing the specialist diagnosis from 43.6% (CI: 0.429; 0.435) at 800€ to 36.9% (CI: 0.361; 0.378) at 50€. To summarize, there is a substantial shift from individuals selecting the diagnostic test to choosing the specialist practitioner with increasing price in the form of a co-pay.

To test hypothesis three, if an increased symptom level does not impact individuals' choice of diagnostic pathway and therefore the test uptake, we can again take a descriptive look and look at the coefficient in our preferred model (5) first. It is important to note that the referred-to symptom level is the description of either *medium-low* or *medium-high* symptoms as described in section 3.3 which participants get randomly allocated to and which stays fixed for all their trade-offs, hence being an exogenous variation between subjects.

Exemplary, we look at the trade-offs where individuals would otherwise have no diagnosis whatsoever, hence the alternatives of diagnostic tests at 200€ and 800€ price points (see table 4.4). The preferences between the different symptom levels are rather similar.

Interpreting the coefficients for the symptom level in model 5 in table 4.3 (-0.0995 for the choice of the diagnostic test, 0.0648 for the choice of the specialist doctor), neither is significant, with the confidence interval stretching from the positive to the negative realm. This shows that

Table 4.4: Respective share of patients choosing no diagnosis instead of the stated alternative

Alternative	Relative (%)	Absolute
Specialist at 80% treatment accuracy		
All	2.33	7
low	2.01	3
high	2.65	4
Specialist at 100% treatment accuracy		
All	1.33	4
low	0.67	1
high	1.99	3
Diagnostic test at 200€ co-pay		
All	10.67	32
low	11.41	17
high	9.93	15
Diagnostic test at 800€ co-pay		
All	35.33	106
low	38.93	58
high	31.79	48
Sample size		300

the symptom level participants were randomly allocated to does not have a significant effect on the chosen diagnostic pathway. Testing, if the interaction of the symptom level and the respective coefficients are different, is not possible due to the computational cost brought by these interaction factors, a common problem of PMLM (Train, 2001). Hence, we cannot reject a null hypothesis of equal estimators to confirm differing estimators. Hence we remain at the observation, that the effects are not significant at p = 0.868 for the diagnostic test and p = 0.909 for the specialist doctor.

On this note, we point out that, while the hypothetical symptom level is not significant, the actual symptom level is significant. Taking a descriptive look, there is no substantial difference between individuals with and without actual symptoms choosing to avoid information when given the alternative of the specialist practitioner, of which there are also very few subjects that make this choice in general. There is a substantially higher share that chooses no diagnosis over the diagnostic test among the subjects with actual symptoms compared to the subjects without. However, when forced to choose between the diagnostic test and the specialist at different treatment accuracy levels, a higher share of the individuals with symptoms chooses the diagnostic test than individuals without symptoms. The same pattern is visible when looking at the average marginal effects of the variable, with all effects being significant at 5%, ceteris paribus. However, the effect size is small in either direction, being comparable with the effect of about 3% additional or lesser treatment accuracy in the most extreme case. As the effect size is very small and this is not part of our hypotheses, we will not discuss this in more depth.

4.3 Information Avoidance and Ostrich Effect

The ostrich effect describes information avoidance given an initial signal about the information. This is the case in our experiment, as the participants get an initial signal, the symptom description (in place of real symptoms one would experience in real life), about a possible negative outcome. To answer hypothesis four, if the ostrich effect is a relevant decision driver, we look at the stated choices. In table 4.4 one can see that only very few individuals choose to avoid going to the doctor, wherein a few more do so when the doctor is less accurate. The share of people not seeking information substantially grows in the trade-offs with the diagnostic test. This is not a surprise as the price associated with the diagnostic test is a major decision driver as analyzed above.

However, as information avoidance is not the main topic of interest in our research, our design is not conceptualized for unambiguous identification of it. There may very well be information avoidance present in the choice preferences of the subjects, which does not show due to the high opposing cost of no diagnosis and hence no treatment whatsoever. In fact, this is likely given the substantial presence of information avoidance in medical decision-making (Golman et al., 2017). Possibly, patients would prefer not knowing but still getting treated. Especially in the case of choosing no diagnosis over the diagnostic test, it may be that individuals are not only driven away by the price but also by the factual knowledge they must face if choosing the test.

Hence, while we cannot conclude that there is no presence of information avoidance whatsoever, we can conclude that the ostrich effect does not drive the decision for patients with EMS symptoms in the choice setting of our experiment.

To answer hypothesis five, if the subject's choice to avoid diagnosis is driven by anxiety or healthcare attitudes, we first look into our preferred model (5). The fact that the coefficients for anxiety, low perceived need, and discomfort are not significant at the 5% level already indicates that there may be no effect. To analyze the effect of anxiety, we derive the average marginal effects of anxiety on the choice of diagnostic options for the subset of choice tasks that portray information avoidance, hence the subset of choice tasks including the option "No diagnosis". We do the same for the two healthcare attitudes, namely discomfort connected to doctor visits and low perceived need for doctor visits. None of the average marginal effects are significant at the 5% level. Hence we cannot reject the null hypothesis that there is no correlation between information avoidance and anxiety, discomfort, or a low perceived need related to doctors' visits, ceteris paribus. It is important, however, to mention that it is very likely, that the small total number of people choosing no diagnosis does not allow for significant data. As discussed for hypothesis 5, less than 3% choose no diagnosis in trade-offs with diagnosis through a specialist. Concluding, we can say that, in the given experimental setting, anxiety, discomfort, and low perceived need do not drive subjects toward seeking a diagnosis in the vast majority of cases. It remains unclear if the discussed factors impact the choice between the diagnostic test and the specialist practitioner. As will be discussed in detail in hypothesis eight, in cases of pure uncertainty resolution people do not name these aspects as decision drivers.

4.4 Uncertainty Resolution

Before analyzing the effects of Uncertainty Resolution we emphasize again that with the decision to display realistic choice options with meaningful descriptions we limit the clear identification of uncertainty resolution. Participants might choose the diagnostic test over the specialist doctor with the same treatment accuracy, time investment, and waiting time until the diagnosis for other, non-UR, related reasons. E.g., one may perceive an ultrasound as more uncomfortable than a swab in the nose or throat. However, we suspect these preferences for the test or against a specialist practitioner to be minor in comparison to the 50€ to 800€ cost and the motivation of UR. Hence we prioritize displaying realistic choice options to meet ethical requirements and to enable relatability for participants. Thereby we also collect data that are more meaningful as they are more likely to portray real choices. To reinforce the validity we control for self-reported preference for UR and rating of reasons that motivate the individual to UR. However, if both factors are low (no preference for UR and no motivation through the named reasons) it may be that the choice is indeed not driven by UR. This also means that, if one questions the assumptions, including when controlling for self-reported preferences, the following conclusions can only be seen as associations that may or may not be driven to an unidentifiable degree by the preference for UR.

Our sixth hypothesis is that individuals seek UR, hence information on their health that brings no direct utility benefit. As argued we assume UR to be the key choice driver for the diagnostic test and factors such as disliking ultrasonic scans to be minor considering the cost point. Further, the model shows a positive effect (0.7866) of the attitude of self-reported preference for UR on the choice of the diagnostic test ($p<0.01$). However, the coefficient magnitude should not be interpreted.

To understand the effect better we isolate the choice sets that exclusively regard UR respectively (hence comparing the diagnostic test to the specialist practitioner with 100% treatment accuracy). In fig. 4.2 we show the share of individuals choosing to resolve uncertainty depending on the hypothetical symptom level and the co-pay for the diagnostic test. We can see that there is a substantial number of individuals that chooses the diagnostic test despite no direct benefit in terms of health outcome. In total 57 out of 300 (19%) resolve uncertainty at a price point of 800€ for the diagnostic test and 184 (61%) out of 300 at a price point of 50€. If we control for a self-reported preference for UR, i.e., individuals answering on the scale of "I would only pay for the diagnostic test if it improves the probability of correct treatment" to "I would pay for the diagnostic test even if it would not change the probability of correct treatment" a minimum of 2 out of 5 (a total of 178 individuals states this), 67% (120 individuals) are willing to pay 50€, and 24% (43) are willing to pay 800€. Setting the bar higher, controlling for a strong self-proclaimed preference for UR, i.e., rating the named question at least 4 out of 5 (a total of 86 individuals state this), 81% (70) are willing to pay 50€ and 38% (33) of the sample are willing to pay 800€. Consequently, even if other reasons play a role in participants' choice for the diagnostic test, there is a substantial number of individuals that have both a strong self-proclaimed preference and a stated preference for UR. As one can see, the share of individuals that are willing to pay 50€ increases from 61% to 81%, and the share that is willing to pay 800€ increases from 19% to

38% when controlling for self-reported UR. We argue, that it is very unlikely that an individual with this combination of stated and self-reported preference for UR is majorly driven by other factors when choosing the diagnostic test. To confirm this we asked the participants to score and state reasons to pay for the resolution of uncertainty given no benefit in treatment and did not receive answers that would indicate other major choice drivers. The exact results will be discussed in detail in section 4.4.1.

The predicted margins of increasing self-reported preference for UR from "no preference for UR" (rating 1/5) to a very strong preference (rating 5/5) increase the share of individuals that choose the diagnostic test from 39.6% (CI: 0.387; 0.405) of individuals to 47.4% (CI: 0.459; 0.489), at mean levels for all other coefficients. Accordingly, there is a small decrease in the share of individuals choosing no diagnosis falling from 18.3% (CI: 0.179; 0.187) to 17.7% (CI: 0.170; 0.183), and a substantial decrease in the percentage of individuals choosing the specialist, dropping from 42% (CI: 0.412; 0.429) to 34.9% (CI: 0.335; 0.363). These marginal effects are all significant at p<0.001. This demonstrates that a self-proclaimed preference for UR is associated with choosing the diagnostic test, and mainly drives individuals seeking a suspicion diagnosis with a specialist to seek a confirmed diagnosis through a diagnostic test.

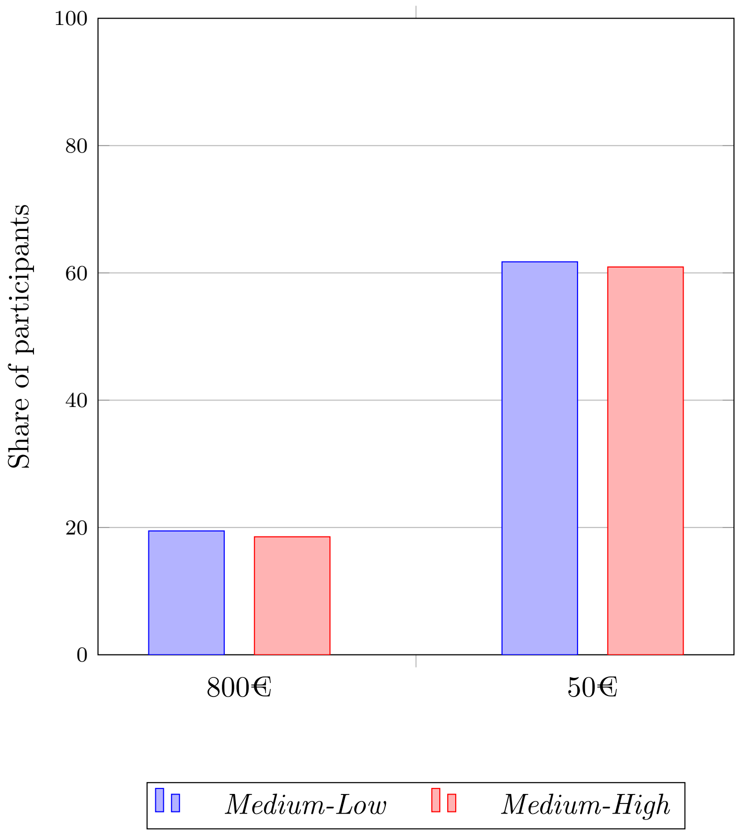


Figure 4.2: Stated preferences for uncertainty resolution in the DCE trade-offs by co-pay and hypothetical symptom level in relative numbers.

4.4.1 Choice drivers

Level of symptoms To understand the effect of the hypothetical symptom level on UR, we first take a descriptive view of stated choices and then calculate the marginal effects of symptom levels in model 5 exclusively for the trade-offs in which the diagnostic test has no treatment benefits. Again this isolates the trade-off on the aspect of pure UR for cost, under the assumption that UR is the only choice driver for the costly diagnostic test.

In fig. 4.2 one can see the relative share of participants, by symptom level, choosing the diagnostic test over a specialist, given no differences in treatment outcome (i.e. driven by a preference for uncertainty resolution). There appears to be little or no difference, even with a slightly higher number of subjects with *medium-low* symptom levels seeking UR.

The average marginal effects of the symptom level on the choice of diagnostic option in cases of UR for the *medium-high* symptom level is (-)0.016 (CI: (-)0.036; 0.005) at p=0.13 on the choice of the diagnostic test and opposingly 0.016 (CI: (-)0.005; 0.036) on the specialist doctor at p=0.13. Hence the hypothetical, randomly allocated symptom level is not significant at the 5% level for the average individual's choice to resolve uncertainty.

Reasons to seek a diagnostic test To better understand the motivation for choosing the diagnostic test, we asked participants an open question: "Please score if these reasons would make you take the diagnostic test, under a cost, even if it does not improve the probability of correct treatment." As discussed in section 3.4, the options they were given are the two most named reasons elicited in qualitative interviews and research prior to this study, to factually know, and to create social acceptance, as well as any "other" reason with a text box to specify.

21 individuals specified their reasoning for "other" (out of 23 that scored it lager than 1/5), wherein all named reasons are directly related to having factual knowledge or social acceptance. For example, multiple women want to know that the specialist practitioner is not just coincidentally treating the right symptoms based on the suspicion diagnosis but is actually treating EMS, to be sure there is not a different underlying disease that they should be aware of. One woman states, that she would like to know for certain if she is fertile. Another wants to be able to account correctly for possible contraindications (simply said negative interactions between a medication and another condition). Similarly, women state to be scared that others, including or especially doctors, might question the diagnosis as long as it is suspicion based. Combining these qualitative observations with the fact that no woman that did not specify the reason "other" scored it at 5/5, only one scored other reasons 4/5, and one 3/5, it seems reasonable to assume that UR resolution driven through social acceptance or the seeking of secured knowledge is the main driver behind choosing the diagnostic test.

Thus, to answer hypothesis eight, if social validation is the most important reason to choose the diagnostic test, and therefore the key driver for uncertainty resolution, we compare the mean scores individuals give for the different reasons for choosing the test, even if the treatment is equally good when choosing the specialist. Individuals score the reasons on a scale from 1 "irrelevant" to 5 "highly relevant". For the entire subject pool, the means are 3.84 for factual knowledge, 3.19 for social acceptance, and 2.72 for other reasons. Using a one-sample multivariate test on equal means verifies that there is a statistically significant difference between the means of the scores at a $p < 0.001$ ($F(2, 918) = 135.50$). When looking only at individuals with a self-proclaimed preference for uncertainty resolution (again, scoring at least 2 out of 5) the mean scores are more pronounced, with scores of 4.13, 3.49, and 2.84 for factual knowledge, social acceptance, and other reasons respectively. The same pattern is visible when only including individuals with a strong self-proclaimed preference for uncertainty Resolution (scoring at least 4 out of 5). The mean scores for the reasons are then correspondingly 4.52, 3.87, and 2.71. In both cases, the multivariate test of differences in means returns p < 0.001 ($F(2{,}498) = 132.66$, and $F(2{,}278) = 151.45$). These results allow two conclusions. For one, not social validation but the pure preference to know is the primary driver for UR. For another, there seem to be no major reasons for the choice of the diagnostic test except reasons that display a preference for pure UR, e.g. in the form of just wanting to know or social validation.

4.5 Robustness check

Given the complexity of many hypotheses and many different aspects that might impact the analyses, we demonstrate the results of the robustness check for the entire analysis with the strictest possible subject group.

We designed the trade-off choices to realistically display choice options as they can be found in the German healthcare market. Also, as discussed, women typically get diagnosed with EMS before 35. Hence, we exclude all participants that are not insured in Germany or are older than 34. Further, in our original analysis, we allocated all unanswered or "prefer not to say" answers to categories that opposed our hypotheses if possible, or to a neutral state. We now exclude all participants that did not answer questions that are relevant to our hypotheses. If conclusions differ, that would indicate that either age, country of insurance, or self-selecting to keep information private through not answering impact the results, and further investigation would be needed.

We share the relevant results of the PMLM in table A.2. Model 6 includes all covariates (equivalent to model 4) and model 7 has the same covariates as the preferred model (5). In the model with all covariates, we find equal coefficient directions and mostly equal significance. There are a few differences in significance. While the effect for treatment accuracy is significant at 1% in the full sample (p=0.000), it is significant at 5% in the robustness test (p=0.020). While the self-reported preference for UR is significant for the choice of the diagnostic test at 1% (p=0.001) in the full sample, it is significant at 5% (p=0.011) in the robust sample. While reporting to have been diagnosed with EMS is positive but insignificant at 5% for the choice of a specialist diagnosis in the full sample (p=.100), it is positive and significant at 5% (p=0.039) for the choice of a diagnosis by the specialist practitioner in the robust sample. Comparing model 7 to the preferred model 5 we find similar shifts. The estimate of treatment accuracy is significant at 1% (p=0.0010), and the self-reported preference for UR in choosing the diagnostic test is again significant at 5% (p=0.012) instead of at 1%. And again reporting an EMS diagnosis has a positive effect at 5% (p=0.039) for the robust sample in the preferred model, unlike the full sample. These results indicate the validity of the full sample, as it is to be expected that significance levels weaken under a smaller sample size. The fact that an EMS diagnosis has a significant effect on the robust sample may indicate that younger women with EMS, likely having been diagnosed recently by a specialist, have a preference for seeking a diagnosis through a specialist, compared to no diagnosis at all.

To ensure the validity of our main analysis we will perform the entire analysis with model 7, the same model but with the robust sample, and report the results in a concise form. For

Hypothesis one, we can confirm the observed effect that a majority of individuals choose the diagnostic test at a price point above 0€, given no benefit in treatment outcome. In the DCE, at a cost (co-pay) of 800€ (100%) for the diagnostic test 24 individuals, or 15.2% choose to resolve uncertainty. This share increases to 92 individuals or 58.2% at a cost of 50€. Again, we can reject the null hypothesis that this is equal to 50% with $t(157) = -12.1522$, $p < 0.001$.

Hypothesis two, testing if the diagnostic test is chosen more often with decreasing price, again is confirmed by the robust sample. The predictive margins, at the price points of 800€ and 50€ respectively, increase for the choice of the diagnostic test from 36.6% (CI: 0.355; 0.378) to 45.2% (CI: 0.439; 0.466), and decrease for the choice of the specialist practitioner from 44.2% (CI:0.431; 0.453) to 36.3% (CI: 0.350; 0.376). These values are very similar to those of the full data set and also significant at 1%. Hence confirming that the diagnostic test is chosen more often with decreasing price.

As in the main results, the average marginal effects for the hypothetical symptom level are not significant at 5%. Hence we cannot find evidence that there is a significant effect of the symptom level. However, we cannot compute the interaction terms of the level with the coefficients to reject the null hypothesis that the coefficients differ, due to very high computational cost. Consequently, we can also not unambiguously confirm hypothesis three, even though we do not find significant effects of the symptom level.

Hypothesis four, if there is no significant occurrence of the ostrich effect, comes to the same conclusion. Only two out of 158 individuals avoid information when trading off no diagnosis with a specialist practitioner treating at 100% accuracy. The largest number opt for no diagnosis when faced with choosing the alternative of the diagnostic test at 800€. Then 59 or 37.4 percent choose no diagnosis. The full results can be found in the table in table A.3 in the appendix. Hence, we equally conclude that in this choice setting the ostrich effect does not drive the decision for patients with EMS symptoms.

Given that the analysis of hypothesis five already suffers from the small number of subjects choosing no diagnosis, it is logical that with the robust sample there is also no significant average marginal effect of healthcare attitudes (anxiety, low perceived need, and discomfort), ceteris paribus. Hence we cannot find a significant effect of healthcare attitudes on the avoidance of diagnostic results.

Looking at the three hypotheses related to uncertainty resolution, we can confirm all results found in the full sample. Firstly (hypothesis six), we find similar descriptive effects. The share of individuals that is willing to pay 50 € and 800€, increases from 58% to 74% and from 15% to 35% respectively, when they state a strong preference for UR in comparison to the full sample. The predicted share (predicted margins) of individuals that choose the diagnostic test increases from 38.0% to 49.5%, when their self-proclaimed preference for uncertainty resolution increases from non (1/5) to very strong (5/5), ceteris paribus. The effect comes with a decrease in the choice of the specialist practitioner from 43.0% to 31.9%, and no change in the choice of no diagnosis. Again the effects are significant at the 1% level. Secondly (hypothesis seven), we do not find significant effects of the hypothetical symptom level on the preference for UR. The average marginal effects of the symptom level are not significant at 5% (p=14.4) on the diagnostic choice in trad-offs without treatment benefit.

Lastly, in hypothesis eight we want to better understand the reasons for choosing the diagnostic test, thereby strengthening the assumption of UR being the main driver. As in the larger sample, we find the reason for factual knowledge to be the most important reason with a mean rating of 3.75 on a scale from 1 (irrelevant) to 5 (highly relevant). This is followed by social acceptance with a mean of 3.17 and other reasons with a mean of 3. A similar pattern is seen in the stricter groups with slight and strong preferences for UR. This reinforces the results of the original data set that not social validation but factual knowledge is the key driver for uncertainty resolution. The multivariate tests testing differences in means of the reasons for the three groups respectively mentioned in the original analysis, namely the entire sample, individuals with a slight preference for UR, and individuals with a strong preference for UR, all confirm the results with $p < 0.001$ ($F(2,528) = 82.13$, $F(2,338) = 142.30$, and $F(2,218) = 153.94$ respectively).

Chapter 5

Discussion and Conclusion

Looking at behavioral health economics research, while information avoidance behavior is thoroughly researched (Galai & Sade, 2003; Golman et al., 2017; Karlsson et al., 2009; Kőszegi, 2003, 2004, 2006; Kőszegi & Rabin, 2008; Li et al., 2021; Panidi, 2015; Schweizer & Szech, 2018), the preference for seeking of information that does not add direct value, i.e. the resolution of uncertainty, is still not well understood. Thus far, there is first evidence regarding back pain, which, however, fails to clearly separate UR from health outcomes (Lillrank, 2003). We combine this gap with one on the applied side of healthcare decision-making. There is a substantial delay in time to diagnosis for EMS that is unparalleled for diseases of similar prevalence (ESHRE GDG, 2022). This leads to substantial costs for individuals, healthcare systems, and society as a whole (Bianconi et al., 2007; Nnoaham et al., 2011; Soliman et al., 2019). We bring these two gaps together by leveraging the unique diagnostic situation for EMS. Following the latest guidelines a treatment decision can be made with the highest accuracy by a specialist practitioner using imaging technology, given a sufficient skill level and equipment of the doctor. Similarly, a novel diagnostic test, available at a cost, can consistently lead to the highest treatment outcomes. However, unlike the practitioner the diagnostic test can factually confirm the diagnosis, thereby enabling resolution of uncertainty.

We leverage this unique decision situation where UR comes at a cost. In this study, we aimed to understand the decision-making of patients at risk of EMS when choosing a diagnostic option. We focus on symptom levels, healthcare attitudes, and willingness to pay for treatment accuracy as well as factual knowledge. Seeking the latter, factual knowledge that does not bring direct benefits in treatment and thus health outcomes, was the core of our research, as this behavior portrays the act of UR. By doing so, we contribute to closing the research gap.

Using a DCE to elicit stated preferences in trade-offs of diagnostic pathways we elicited preferences regarding the willingness to pay for treatment accuracy and UR while testing for the effect of the symptom level by randomly allocating participants into two groups with higher and lower symptom scenarios on which they based their choices. We inquired about their demographic profile, healthcare attitude, health status, and motivations for choosing the uncertainty-resolving diagnostic test over a specialist practitioner. We found choice behavior that deviates from traditional economic theory, as subjects valued not only treatment accuracy but also seek certain information that does not improve treatment outcomes. We did not find information avoidance to be a substantial choice driver, unlike in many other medical settings (Golman et al., 2017). As discussed, this might be due to the experience a patient has. Feeling healthy despite having cancer might incentivize ignoring the illness to continue living happily. On the other hand, when suffering from EMS symptoms, there is not much hedonic value for the patient in avoiding a diagnosis.

Further, we did not find evidence that the hypothetical symptom level affected the choice. One might expect an increased symptom level to increase willingness to pay, as more symptoms are directly associated with a higher risk of having EMS (Ballard et al., 2008). However, our results are in line with findings from Li et al. (2021), on which we built our hypothesis, who find individuals to self-select out of a diagnostic test with increasing price, independent of their risk level. Nonetheless, one must keep in mind that these results may also be caused by the hypothetical decision situation in which the participating women may have struggled to directly adapt their choices to the described symptoms.

Carefully balancing ethical requirements, methodological concerns, and real-world applicability, we opted for a design that does not allow us to unequivocally isolate the preference for UR. However, given the design and the comparatively insignificant other benefits of choosing a diagnostic test over a specialist practitioner's diagnosis with the same treatment outcome, we assume that the choice is primarily driven by the preference for UR. To back this assumption we control for healthcare attitudes related to avoiding medical professionals, ask participants to explicitly state their preference for UR, and to name reasons why they would choose the diagnostic test over the specialist practitioner given no treatment benefit. The implemented control measures do not give any indication of another driver for the choice apart from a preference for UR, hence leading us to the conclusion that UR is indeed a key choice driver in the choice of a diagnostic pathway for women with EMS risk.

However, the knowledge on UR in medical treatment is still very limited. To better understand possible reasons why individuals may seek a resolution of uncertainty, which has not been researched much, it may be helpful to look at the counteraction, information avoidance. Golman et al. (2017) categorize existing literature into two main reasons for information avoidance with many sub-categories. Hedonic considerations (desiring to avoid negative information) and strategic considerations (e.g., as a type of commitment device, as individuals may expect the information may worsen their situation, their motivation, or bargaining power). Many concepts of the latter may seem less relevant in this specific circumstance. Contrarily one might say not the avoidance of information but factual knowledge of an EMS diagnosis could act as a commitment device to oneself (committing to the treatment which may have side-effects (hormone treatment) or may come with risk and discomfort (surgery)) or to force others to commit to, e.g., more empathetic behavior, as a factual diagnosis, may give less room to question the patient's situation (thus the counteract of interpersonal strategic avoidance, wherein information gets avoided as a strategy to manipulate another's actions). It is important to reinforce that from a medical point of view the diagnosis and the treatment decision should not be questioned solely on the basis of the diagnostic pathway.

In summary, we find first evidence for UR and its importance in patient decisions. Indeed, Golman et al. (2017) argue hedonic consequences, of not only information avoidance but also information acquisition (hence, uncertainty resolution), should be part of any welfare calculation. Caplin and Leahy (2004), Lipnowski and Mathevet (2017) and Rothenhäusler et al. (2013) recognize the role information may have that exceeds the direct treatment in proposing mechanisms for the optimal provision of information to patients with anxiety. With an increased understanding of UR, it could be similarly accounted for in the future, thereby improving choice architecture, healthcare systems, and potentially health outcomes. This could contribute to a substantial reduction in diagnostic delays and in healthcare costs.

5.1 Limitations

Looking at the participating population and the distribution channels, we achieved to spread the questionnaire into many different groups. Nonetheless, the sample selection is biased toward educated people and one must be aware that the non-random, non-obligatory participation leads to certain participants self-selecting into the sample, as well as possibly certain participants self-selecting out of the survey before starting it or during the process. 75 out of 375 participants self-selected out of the questionnaire, for reasons that cannot be elicited. These may be random, such as being interrupted, or non-random, such as lack of interest, understanding, or ability to relate with EMS patients. This is visible in the over population-average share of participants that self-report an EMS diagnosis. This could be through individuals with EMS being less likely to drop out of the survey before participation, or being more likely to take part in general, if people mentioned the topic of the survey when they shared it in their social circles. Hence, the sample suffers from a selection bias that cannot be completely eliminated. This is directly connected to the fact that the situation is hypothetical, as we performed the DCE in an online survey. This is always to be taken into account when interpreting the results. Different research has proven the validity of such settings, also when using specific subject pools such as students. However, it is important to note that the exact effect size often is not meaningful when translating the results into the real world or to different population groups. Nonetheless, the qualitative effects hold (Brosig-Koch et al., 2016; Stoop, 2012; Vissers et al., 2001).

A key limitation is the use of labeled alternatives. As discussed using labels has many benefits strengthening the results for real-world applicability and enabling the MLM estimator. Nonetheless, this decision also entails substantial downsides. Giving a name to options, which may make people consider other aspects, such as the discomfort of a swab in the test or ultrasonic imaging, prohibits clear, unambiguous identification of UR. Another key limitation is the exclusive use of attribute combinations that can be found in the real decision situations of the German healthcare market. There are good reasons for this, as creating non-realistic attribute combinations may be unethical, creating wrong pictures in patients' minds. However, this has consequences. For example, not varying the price for all alternatives (i.e. also giving a price for some cases of no diagnosis and the specialist diagnosis) decreases the quality of trade-offs and thus increases the number of participants needed for accurate estimation. Again this lack of certain attribute combinations impacts the clear identification of UR. While we do argue that the benefits are worth these downsides and use multiple measures to ensure there are no other effects driving the decision, we cannot unambiguously separate UR from other factors to identify a clear causal effect.

Further, one might argue, given that the design of the experiment supplies substantial information about EMS, the setting creates the mindset of a patient that thinks they have EMS, even though, or especially because an approximate likelihood (one-third) is shared with them. This could impact the prevalence of information-avoiding behavior or explain why there is a small/no difference between symptom levels. Research on this is mixed. Nosarti et al. (2000) show that patients who think they have cancer delayed their doctor's visit less. Li et al. (2021) found the uptake of a free diabetes test to increase with subjective risk. On the other hand, patients that know they have a high risk (e.g. through genetics, this is objective, not subjective risk) tend to avoid doctor visits (Golman et al., 2017). It is key to differentiate the subjective and objective risk - one may subjectively think one has cancer while having absolutely no objective risk. As we stated the objective risk-probability subjective factors should, theoretically, not play a role. This discussed framing through the supplied information may impact the decision-making, hence, the results regarding information avoidance have to be interpreted with this in mind. Given that the information provided is concise and objective, a divergence in the real world or other framing would highlight the importance of information architecture. This leads us to the need for, and the potential of, future research.

5.2 Future Research

As this DCE entered a novel field of research in both understanding the diagnostic choice behavior of patients with EMS risk and the topic of UR in healthcare, there is a large realm of potential future research. The potential impact is substantial as improving the understanding of choices that patients with EMS risk make has substantial potential in improving the design of healthcare systems, regulation, and information architecture.

One obvious continuation to better understand EMS patients is replicating the study in a real-world setting, e.g. with patients that are actually in this decision situation at a primary care physician's office or that seek advice in professional online tools for EMS risk assessment such as the "Endo AI", an online self-test for EMS risk which is certified as medical device (Rohloff, 2023).

The fact that very few patients avoid diagnoses potentially indicates that research to improve time-to-diagnosis should focus on other aspects causing diagnostic delay such as awareness and quality of information on the patient as well as the practitioner side, and quality of diagnostic procedures. However, it is to be established if the observed behavior holds in situations that do not include the framing given in this experiment. If the framing would impact the decision-making significantly this may also indicate that the information architecture plays a key role.

To expand the understanding of UR one could replicate the study with different diseases and thereby understand how the potential illness impacts the preference. Additionally, deeper research on UR is needed to understand why people have this preference, how it precisely affects their choices, and if it is beneficial to shift people away from this behavior, or how to account for and potentially fulfill these preferences. When striving to finally improve health outcomes, it is paramount to research the exact reasons or dynamics of behavior. For example, if people do want to resolve uncertainty, what are the reasons in real life? Thereby, one can assess if the UR through the diagnostic test is crucial and can actually increase health outcomes, e.g. through feeling better psychologically, or if it makes sense to change information architecture to achieve similar effects of psychological safety through the specialist's diagnosis.

5.3 Conclusion

The delay in diagnoses and the underdiagnosing of EMS has substantial costs for society with over 100 million euros per year for the German healthcare system alone. The quality of life reduction for patients and their social circles is estimated to impact over 190 million women every day. Not directly connected to this, UR may have consequences similar to hedonic information avoidance that impact the patient's utility function and should then also be taken into account to optimize aspects like information architecture to achieve the best possible outcomes in the patient's own interest. However, research on UR in healthcare is lacking. Bringing the picture together in this work, we simultaneously address some aspects of the problem of severe underdiagnosis of EMS as well as make a first step to understand the role of UR in healthcare settings.

In our discrete choice experiment, we do not find the symptom level of patients, nor a preference for information avoidance to drive the choices of subjects with EMS risk. However, they show a clear preference for taking a diagnostic test and thereby resolving uncertainty. We inquire for and thus rule out other drivers of this choice to the best possible degree to show the significant role of UR in diagnostic pathway decisions of patients with EMS risk. We humbly hope that these findings can inform and motivate future research and consequentially a better design of the healthcare systems, and the information architecture of EMS treatment.

Bibliography

Adamson, G. D., Kennedy, S., & Hummelshoj, L. (2010). Creating Solutions in Endometriosis: Global Collaboration through the World Endometriosis Research Foundation. *Journal of Endometriosis*, *2*(1), 3–6. `https://doi.org/10.1177/228402651000200102`

Ballard, K. D., Seaman, H. E., De Vries, C. S., & Wright, J. T. (2008). Can symptomatology help in the diagnosis of endometriosis? Findings from a national case-control study - Part 1. *BJOG: An International Journal of Obstetrics and Gynaecology*, *115*(11), 1382–1391. `https://doi.org/10.1111/j.1471-0528.2008.01878.x`

Benartzi, S., & Thaler, R. H. (1995). Myopic Loss Aversion and the Equity Premium Puzzle [Publisher: Oxford University Press]. *The Quarterly Journal of Economics*, *110*(1), 73–92. `https://doi.org/10.2307/2118511`

Bendifallah, S., Dabi, Y., Suisse, S., Delbos, L., Poilblanc, M., Descamps, P., Golfier, F., Jornea, L., Bouteiller, D., Touboul, C., Puchar, A., & Daraï, E. (2022). Endometriosis Associated-miRNome Analysis of Blood Samples: A Prospective Study. *Diagnostics (Basel, Switzerland)*, *12*(5), 1150. `https://doi.org/10.3390/diagnostics12051150`

Bendifallah, S., Suisse, S., Puchar, A., Delbos, L., Poilblanc, M., Descamps, P., Golfier, F., Jornea, L., Bouteiller, D., Touboul, C., Dabi, Y., & Daraï, E. (2022). Salivary MicroRNA Signature for Diagnosis of Endometriosis. *Journal of Clinical Medicine*, *11*(3), 612. `https://doi.org/10.3390/jcm11030612`

Berkley, K. J., Rapkin, A. J., & Papka, R. E. (2005). The Pains of Endometriosis. *Science*, *308*(5728), 1587–1589. `https://doi.org/10.1126/science.1111445`

Bianconi, L., Hummelshoj, L., Coccia, M. E., Vigano, P., Vittori, G., Veit, J., Music, R., Tomassini, A., & D'Hooghe, T. (2007). Recognizing endometriosis as a social disease: The European Union-encouraged Italian Senate approach. *Fertility and Sterility*, *88*(5), 1285–1287. `https://doi.org/10.1016/j.fertnstert.2007.07.1324`

Bista, S., Yu, R., Shete, S., & Shastri, S. (2021). Factors Associated with Healthcare Avoidance among US Adults: Analyses of the 2019 Healthy Information National Trends Survey. *Journal of Family Medicine*, *8*(9). `https://doi.org/10.26420/jfammed.2021.1280`

Bowen, D., McTiernan, A., Burke, W., Powers, D., Pruski, J., Durfy, S., Gralow, J., & Malone, K. (1999). Participation in breast cancer risk counseling among women with a family history. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *8*(7), 581–585.

BPB. (2022). Krankenversicherungsschutz. Retrieved June 7, 2023, from https://www.bpb.de/kurz-knapp/zahlen-und-fakten/soziale-situation-in-deutschland/61813/krankenversicherungsschutz/

Brooks, R., & De Charro, F. (2003). *The measurement and valuation of health status using EQ-5D: A European perspective; Summary and future EQ-5D work [Chapter 14]*. https://eq-5dpublications.euroqol.org/details?id=0_54025&nosearchform=false

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & Wiesen, D. (2016). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. *Journal of Economic Behavior & Organization*, *131*, 17–23. https://doi.org/10.1016/j.jebo.2015.04.011

Caplin, A., & Leahy, J. (2004). The supply of information by a concerned expert [Publisher: Oxford University Press / USA]. *Economic Journal*, *114*(497), 487–505. https://doi.org/10.1111/j.0013-0133.2004.0228a.x

Ceylan, B., & Özerdoğan, N. (2015). Factors affecting age of onset of menopause and determination of quality of life in menopause. *Turkish Journal of Obstetrics and Gynecology*, *12*(1), 43–49. https://doi.org/10.4274/tjod.79836

Chew, H., & Ho, J. L. (1994). Hope: An empirical study of attitude toward the timing of uncertainty resolution.

De Bekker-Grob, E. (2009). *Discrete Choice Experiments in Health Care: Theory and applications*. Retrieved July 16, 2023, from https://repub.eur.nl/pub/21908/

De Bekker-Grob, E., Donkers, B., Jonker, M. F., & Stolk, E. A. (2015). Sample Size Requirements for Discrete-Choice Experiments in Healthcare: A Practical Guide. *The Patient - Patient-Centered Outcomes Research*, *8*(5), 373–384. https://doi.org/10.1007/s40271-015-0118-z

de Bekker-Grob, E. W., Donkers, B., Jonker, M. F., & Stolk, E. A. (2015). Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. *Patient*, *8*(5), 373–384. https://doi.org/10.1007/s40271-015-0118-z

Derrett, S., Davie G & Samaranayaka A. (2016). *Understanding relationships between the EQ-5D and Personal Wellbeing*. https://eq-5dpublications.euroqol.org/download?id=0_56618&fileId=56717

Eluthia. (2023). Endotest – Eluthia. Retrieved June 15, 2023, from https://www.eluthia.com/endotest/

ESHRE GDG. (2022). *Endometriosis Guideline of European Society of Human Reproduction and Embryology* (tech. rep.). www.eshre.eu/guidelines

Galai, D., & Sade, O. (2003). The "Ostrich Effect" and the relationship between the liquidity and the yields of financial assets. *The Journal of Business, Vol. 79, No. 5, pp. 2741-2759*. https://doi.org/https://doi.org/10.1086/505250

Garrido, A. Q., Wainstein, A. J. A., Brandão, M. P. A., de Vasconcellos Santos, F. A., Bittencourt, F. V., Ledsham, C., & Drummond-Lage, A. P. (2020). Diagnosis of Cutaneous Melanoma: The Gap Between the Knowledge of General Practitioners and Dermatologists in a Brazilian Population. *Journal of Cancer Education: The Official Journal of* *the American Association for Cancer Education*, *35*(4), 819–825. https://doi.org/10.1007/s13187-020-01735-z

Goehring, J., Drewes, M., Kalder, M., & Kostev, K. (2022). Germany Endometriosis Pattern Changes; Prevalence and Therapy over 2010 and 2019 Years: A Retrospective Cross Sectional Study. *International Journal of Fertility and Sterility*, *16*(2). https://doi.org/10.22074/IJFS.2021.528397.1113

Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information Avoidance. *Journal of Economic Literature*, *55*(1), 96–135. https://doi.org/10.1257/jel.20151245

Gusi, N., Olivares, P. R., & Rajendram, R. (2010). The EQ-5D Health-Related Quality of Life Questionnaire. In V. R. Preedy & R. R. Watson (Eds.), *Handbook of Disease Burdens and Quality of Life Measures* (pp. 87–99). Springer. https://doi.org/10.1007/978-0-387-78665-0_5

Harmens, M., Sedgewick, F., & Hobson, H. (2022). The Quest for Acceptance: A Blog-Based Study of Autistic Women's Experiences and Well-Being During Autism Identification and Diagnosis [Publisher: Mary Ann Liebert, Inc., publishers]. *Autism in Adulthood*, *4*(1), 42–51. https://doi.org/10.1089/aut.2021.0016

HAW Hamburg. (2022). *Checkliste Ethische Selbstbewertung - Ethikkommission HAW Hamburg November 2022* (tech. rep.). Ethics commission, HAW Hamburg - Competence Center Gesundheit. https://www.haw-hamburg.de/fileadmin/Forschung/PDF/Ethik/Checkliste_Ethische_Selbsbewertung.docx

Hirst, L. W., & Smith, J. (2020). Accuracy of diagnosis of pterygium by optometrists and general practitioners in Australia. *Clinical & Experimental Optometry*, *103*(2), 197–200. https://doi.org/10.1111/cxo.12916

Hudson, N. (2021). The missed disease? Endometriosis as an example of 'undone science'. *Reproductive Biomedicine & Society Online*, *14*, 20–27. https://doi.org/10.1016/j.rbms.2021.07.003

Illum, L. R. H., Forman, A., Melgaard, A., Hansen, K. E., Hansen, S. N., Nyegaard, M., Hummelshoj, L., & Rytter, D. (2022). Temporal and regional differences in the incidence of hospital-diagnosed endometriosis: A Danish population-based study. *Acta Obstetricia et Gynecologica Scandinavica*, *101*(7), 737–746. https://doi.org/10.1111/aogs.14364

IW. (2023). *Alle Rechner – Arm-und-Reich.de* (tech. rep.). Retrieved June 12, 2023, from https://www.arm-und-reich.de/alle-rechner/

Kahneman, D., Sibony, O., & Sunstein, C. (2021). *Noise: A Flaw in Human Judgment*. Little, Brown. https://books.google.nl/books?id=g2MBEAAAQBAJ

Karlsson, N., Loewenstein, G., & Seppi, D. (2009). The ostrich effect: Selective attention to information [Publisher: Springer]. *Journal of Risk and Uncertainty*, *38*(2), 95–115. Retrieved June 6, 2023, from https://www.jstor.org/stable/41761376

Kash, K. M., Holland, J. C., Halper, M. S., & Miller, D. G. (1992). Psychological Distress and Surveillance Behaviors of Women With a Family History of Breast Cancer. *JNCI: Journal of the National Cancer Institute*, *84*(1), 24–30. https://doi.org/10.1093/jnci/84.1.24

Kőszegi, B. (2003). Health anxiety and patient behavior. *Journal of Health Economics*, *22*(6), 1073–1084. https://doi.org/10.1016/j.jhealeco.2003.06.002

Kőszegi, B. (2004). *Emotional Agency: The case of the doctor-patient relationship* (tech. rep.). Citeseer.

Kőszegi, B. (2006). Emotional Agency [Publisher: Oxford University Press / USA]. *Quarterly Journal of Economics*, *121*(1), 121–155. https://doi.org/10.1162/qjec.2006.121.1.121

Kőszegi, B., & Rabin, M. (2008). Revealed mistakes and revealed preferences. *The foundations of positive and normative economics: a handbook*, 193–209.

Lancsar, E., & Louviere, J. (2008). Conducting Discrete Choice Experiments to Inform Healthcare Decision Making A User's Guide. *PharmacoEconomics*, 661.

Li, Y., Meng, J., Song, C., & Zheng, K. (2021). Information Avoidance and Medical Screening: A Field Experiment in China. *Management Science*, *67*(7), 4252–4272. https://doi.org/10.1287/mnsc.2020.3723

Lillrank, A. (2003). Back pain and the resolution of diagnostic uncertainty in illness narratives. *Social Science & Medicine*, *57*(6), 1045–1054. https://doi.org/10.1016/S0277-9536(02)00479-3

Lipnowski, E., & Mathevet, L. (2017). Disclosure to a Psychological Audience. *American Economic Journal: Microeconomics*, *10*. https://doi.org/10.1257/mic.20160247

Lovallo, D., & Kahneman, D. (2000). Living with uncertainty: Attractiveness and resolution timing. *Journal of Behavioral Decision Making*, *13*(2), 179–190. https://doi.org/10.1002/(SICI)1099-0771(200004/06)13:2<179::AID-BDM332>3.0.CO;2-J

Mant, J., Fitzmaurice, D. A., Hobbs, F. D. R., Jowett, S., Murray, E. T., Holder, R., Davies, M., & Lip, G. Y. H. (2007). Accuracy of diagnosing atrial fibrillation on electrocardiogram by primary care practitioners and interpretative diagnostic software: Analysis of data from screening for atrial fibrillation in the elderly (SAFE) trial. *BMJ (Clinical research ed.)*, *335*(7616), 380. https://doi.org/10.1136/bmj.39227.551713.AE

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, *15*(5), 447–470. https://doi.org/10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1

McFadden et al., D. (1973). Conditional logit analysis of qualitative choice behavior.

Mitchell, A. J., Meader, N., & Pentzek, M. (2011). Clinical recognition of dementia and cognitive impairment in primary care: A meta-analysis of physician accuracy. *Acta Psychiatrica Scandinavica*, *124*(3), 165–183. https://doi.org/10.1111/j.1600-0447.2011.01730.x

Müller, S., Scealy, J. L., & Welsh, A. H. (2013). Model Selection in Linear Mixed Models [arXiv:1306.2427 [stat]]. *Statistical Science*, *28*(2). https://doi.org/10.1214/12-STS410

Nnoaham, K. E., Hummelshoj, L., Webster, P., d'Hooghe, T., De Cicco Nardone, F., De Cicco Nardone, C., Jenkinson, C., Kennedy, S. H., & Zondervan, K. T. (2011). Impact of endometriosis on quality of life and work productivity: A multicenter study across ten countries. *Fertility and Sterility*, *96*(2), 366–373.e8. https://doi.org/10.1016/j.fertnstert.2011.05.090

Nosarti, C., Crayford, T., Roberts, J. V., Elias, E., McKenzie, K., & David, A. S. (2000). Delay in presentation of symptomatic referrals to a breast clinic: Patient and system factors. *British Journal of Cancer*, *82*(3), 742–748. https://doi.org/10.1054/bjoc.1999.0990

OECD. (2022). Bildung auf einen Blick 2022 — READ online. Retrieved June 12, 2023, from https://read.oecd.ilibrary.org/education/bildung-auf-einen-blick-2022_dd19b10a-de

Orme, B. (2019). Chapter 7 sample Size Issues for Conjoint Analysis. In *Getting Started with Conjoint Analysis: Strategies for product Design and Pricing Research. Fourth Edition, Madison, Wis.: Research Publishers LLC*. Retrieved May 31, 2023, from https://sawtoothsoftware.com/resources/technical-papers/sample-size-issues-for-conjoint-analysis-studies

Panidi, K. (2015). Ostrich Effect in Health Care Decisions: Theory and Empirical Evidence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2932181

Pearmain, D., & Kroes, E. P. (1990). *Stated preference techniques : a guide to practice*. Steer Davies & Gleave Ltd. ; Hague Consultancy Group.

Pugsley, Z., & Ballard, K. (2007). Management of endometriosis in general practice: The pathway to diagnosis. *The British journal of general practice : the journal of the Royal College of General Practitioners*, *57*, 470–6.

Qualtrics. (2023). Qualtrics Survey Software. https://www.qualtrics.com

Rabe-Hesketh, S., & Skrondal, A. (2022). Chapter 12: Nominal responses and discrete choice. In S. Rabe-Hesketh & A. Skrondal (Eds.), *Multilevel and longitudinal modeling using stata* (4th). Stata Press.

RatSWD. (2017). *Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften* (tech. rep.). Rat für Sozial- und Wirtschaftsdaten. Berlin. https://doi.org/DOI:10.17620/02671.1

Rohloff, N. (2023). EndoAI - Endometriose-Selbsttest. Retrieved July 13, 2023, from https://endometriose.app/symptome/endometriose-test

Rose, J. M., & Bliemer, M. C. (2009). Constructing efficient stated choice experimental designs. *Transport Reviews*, *29*(5), 587–617. https://doi.org/10.1080/01441640902827623

Rothenhäusler, D., Schweizer, N., & Szech, N. (2013). Institutions, Shared Guilt, and Moral Transgression. *Discussion Paper SP II 2013–305*. Retrieved July 14, 2023, from https://bibliothek.wzb.eu/pdf/2013/ii13-305.pdf

Ryan, M., Bate, A., Eastmond, C. J., & Ludbrook, A. (2001). *Use of discrete choice experiments to elicit preferences* (tech. rep.). www.qualityhealthcare.com

Ryan, M., Gerard, K., & Amaya-Amaya, M. (2008). *Using discrete choice experiments to value health and health care* (Vol. 11). Springer.

Schweizer, N., & Szech, N. (2018). Optimal Revelation of Life-Changing Information. *Management Science*, *64*(11), 5250–5262. https://doi.org/10.1287/mnsc.2017.2913

Soekhai, V., Whichello, C., Levitan, B., Veldwijk, J., Pinto, C. A., Donkers, B., Huys, I., Van Overbeeke, E., Juhaeri, J., & De Bekker-Grob, E. W. (2019). Methods for exploring and eliciting patient preferences in the medical product lifecycle: A literature review. *Drug Discovery Today*, *24*(7), 1324–1331. https://doi.org/10.1016/j.drudis.2019.05.001

Soliman, A. M., Surrey, E. S., Bonafede, M., Nelson, J. K., Vora, J. B., & Agarwal, S. K. (2019). Health Care Utilization and Costs Associated with Endometriosis Among Women with Medicaid Insurance [Publisher: Academy of Managed Care Pharmacy]. *Journal of Managed Care & Specialty Pharmacy*, *25*(5), 566–572. https://doi.org/10.18553/jmcp.2019.25.5.566

Statista. (2023). Studienanfängerquote in Deutschland bis 2022. Retrieved June 12, 2023, from https://de.statista.com/statistik/daten/studie/72005/umfrage/entwicklung-der-studienanfaengerquote/

Stoop, J. (2012). From the lab to the field: Envelopes, dictators and manners [Number: 37048 Publisher: University Library of Munich, Germany]. *MPRA Paper*. Retrieved July 16, 2023, from https://ideas.repec.org//p/pra/mprapa/37048.html

Taber, J. M., Leyva, B., & Persoskie, A. (2015). Why do People Avoid Medical Care? A Qualitative Study Using National Data. *Journal of General Internal Medicine*, *30*(3), 290–297. https://doi.org/10.1007/s11606-014-3089-1

Train, K. E. (2001). *Discrete Choice Methods with Simulation* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511805271

Vissers, G., Heyne, G., Peters, V., & Geurts, J. (2001). The Validity of Laboratory Research in Social and Behavioral Science. *Quality and Quantity*, *35*, 129–145. https://doi.org/10.1023/A:1010319117701

WHO. (2023). Endometriosis. Retrieved June 14, 2023, from https://www.who.int/news-room/fact-sheets/detail/endometriosis

Appendix A

Appendix

A.1 Survey

In this appendix section, we share the survey with the exact visuals that participants faced. We split the survey into different sections, starting with the onboarding.

Diagnostic Choices

This survey aims to understand your decisions as a patient better. Please answer all questions thoughtfully and decide as you would when seeking real treatment. Thank you for your contribution to research.

The following survey is being conducted for research purposes only. Your responses will be kept anonymous and confidential, and will only be used in aggregate form for statistical analysis. Your participation is voluntary, and you may withdraw from the survey at any time without penalty. Please acknowledge that you have read and understood this disclaimer and consent to participate in this research study by choosing "yes".

- Yes
- No

If you have any questions or comments please contact:
Stephan Rothenberger
Erasmus University Rotterdam
sr@student.eur.nl

A.1.1 Demographics

How old are you?

- Under 18
- 18-24 years old
- 25-34 years old
- 35-44 years old
- 45-54 years old
- 55-64 years old
- 65+ years old

How do you describe yourself?

- Male
- Female
- Non-binary / third gender
- Prefer to self-describe
- Prefer not to say

Are you health-insured in Germany?

- Yes, public (statutory) health insurance (SHI)
- Yes, private health insurance (PHI)
- No, I'm insured in

What is the highest level of education you have completed?

- Some Primary Education
- Completed Primary
- Some Secondary Education
- Completed Secondary
- Vocational or Similar
- Some University but no degree
- University Bachelors degree
- Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS etc.)
- Prefer not to say

What was your total household income before taxes during the past 12 months in Euros?

- Less than 25,000 Euros per year
- 25,000 - 49,999 Euros per year
- 50,000 - 99,999 Euros per year
- 100,000 - 199,999 Euros per year
- More than 200,000 Euros per year
- Prefer not to say

A.1.2 Discrete Choice Experiment

The following contains the information every subject received and one trade-off question. Every trade-off question included the description of the hypothetical medium-high or medium-low symptom level that a subject was randomly selected into. Here the example contains the medium-low description.

In the following, you will be repeatedly asked to decide between two diagnostic options for Endometriosis. There are 10 trade-off questions.

Please imagine that you suffer from the symptoms described on the following page and decide as you would in real life. About 1/3 of patients with these symptoms actually have Endometriosis.

Some questions might seem redundant, this is part of the process.

Endometriosis is a chronic condition affecting women of childbearing age, in which tissue similar to the lining of the uterus grows outside of the uterus, potentially causing pain, inflammation, scarring, and infertility.

A new diagnostic test using swabs (like Covid-19 tests) can accurately diagnose Endometriosis, however, it is not covered by health insurance.

Following the latest diagnostic standard, **health insurance covers a suspicion-based diagnosis using imaging technology** (ultrasound or MRI), thereby avoiding any form of surgical intervention. **Accordingly, doctors treat you based only on their suspicion, without a definitive diagnosis.**
The probability of correct treatment, using the suspicion-based diagnosis of a well-informed specialist with modern technology can be as high as the probability of correct treatment based on a diagnostic test. However, the probability may decrease depending on the doctors skill or their availible technology.

In some cases, your treatment is similar even if the diagnosis is faulty or not Endometriosis. Treatment typically entails suppressing symptoms through hormone and pain treatments. However, there is currently no cure and symptoms often recur if therapy is ended.

Consequently, **the only downside of a treatment relying on a suspicion-based diagnosis made by a well-informed specialist is that you do not have a factual confirmation of Endometriosis.** Diagnosis can be an important psychological factor for women suffering from the symptoms of this otherwise invisible disease creating acceptance for themselves and in their social environment/workplace.
Endometriosis is strongly underdiagnosed with an estimated 6-10% of women affected and an average time to diagnosis of 10 years.

Imagine you experience more frequent severe menstrual cramps and pain than your peers, as well as bleeding after intercourse on rare occasions.

Among the following diagnostic options, which one do you prefer?

	Option 1	Option 2
Co-pay / Out-of-Pocket costs	0 Euros (0%)	0 Euros (0%)
Probability of correct treatment	100%	0%
Total time needed for appointment, incl. travel	3 hours	No appointment
Time until diagnosis	21 days	No Result
Diagnostic pathway	**Specialist Doctor (Suspicion-based diagnosis)**	**No diagnosis**

	Option 1	Option 2
Your choice:	○	○

A.1.3 Medical self-reporting, healthcare attitudes and preferences

The following images display the questions that were asked regarding health state, medical attitude and choice preferences.

Have you been diagnosed with Endometriosis by a medical professional?

- Yes ○
- No ○
- Prefer not to say ○

Please indicate which level of severity best describes your **actual** health status.

We remind you that you are not obliged to answer these questions. However, the answers are anonymized and improve the validity of our research.

	No	Slight	Moderate	Severe	Extreme	Prefer not to say
I am anxious / depressed	○	○	○	○	○	○

Please indicate which level of severity best describes your **actual** health status, **taking the below-listed endometriosis symptoms into account**.
We remind you that you are not obliged to answer these questions. However, the answers are anonymized and improve the validity of our research.

- Chronic pelvic pain that worsens before the onset of menstruation
- Infertility
- Menstrual pain that you perceive as abnormal
- Pre- or postmenstrual bleeding
- Pain during intercourse

	No	Slight	Moderate	Severe	Extreme	Prefer not to say
I have __ problems doing my usual activities due to above-described symptoms	○	○	○	○	○	○
I have __ pain/discomfort due to above-described symptoms	○	○	○	○	○	○

Please position yourself on the following scales

I only go to the doctor when really needed, as I try to solve issues myself	○ ○ ○ ○ ○	I tend to go to the doctor quickly when I'm unsure
I am very comfortable with going to the doctor	○ ○ ○ ○ ○	I avoid going to doctors due to discomfort or anxiety
I would only pay for the diagnostic test if it improves the probability of correct treatment	○ ○ ○ ○ ○	I would pay for the diagnostic test even if it would not change the probability of correct treatment, e.g. to factually know if I have Endometriosis

Please score if these reasons would make you take the diagnostic test, under a cost, even if it does not improve the probability of correct treatment.

1 = irrelevant; 5 = Highly relevant

1 2 3 4 5

To factually know the diagnosis

1

To create social acceptance for myself/in my environment

2

Other (please specify):

3

A.1.4 Outro

Thank you for your time, go to the next page to finalize the survey.

If you have any questions or comments please contact:

Stephan Rothenberger

Erasmus University Rotterdam

sr@student.eur.nl

If you wish to learn more about Endometriosis visit the page of the WHO (World Health Organization).

Centers with highest diagnostic standards in Europe can be found on the page of the "Endometriosis Research Foundation".

Figure A.1: Outro message for participants

To guarantee the highest possible quality of results only (cisgender) female individuals can take part in this study, due to research limitations, the nature of Endometriosis, its diagnosis, and its treatment.
Further, if you are below 18, you cannot take part due to legal reasons.
We appreciate your understanding.

We kindly ask you to share this study with your social circle to help us understand female health better.

Link

If you have any questions, concerns, or comments please contact:

Stephan Rothenberger

Erasmus University Rotterdam

sr@student.eur.nl

Figure A.2: Outro message for filtered participants that cannot take part in the survey.

A.2 Additional information on results

Reasons for uncertainty resolution In A.1 we give an overview of the average scores individuals give for the different reasons for choosing the test, even if the treatment is equally good when choosing the specialist. Thereby we give a transparent view of different subgroups and the respective number of observations in these groups. We provide the average scores for the reasons by subgroup. The general overview (1) shows the scoring of all participants. The self-proclaimed preferences for uncertainty resolution shows how people scored the reasons if they positioned themselves on a scale of *"I would only pay for the diagnostic test if it improves the probability of correct treatment"* to *"I would pay for the diagnostic test even if it would not change the probability of correct treatment"* larger than 1, indicating a slight preference for uncertainty resolution (2), or at least 4, indicating a strong preference for uncertainty resolution (3). Stated preference, ie., people that chose to resolve uncertainty in the choice tasks are given for two levels of price points, 50€ (4) and 800€ (5) respectively. (6) to (9) are combined measures of the before mentioned. All sample sizes are reported in parentheses. These vary in groups as well, as responses to these scoring questions were not obligatory, as forcing people to pinpoint their motivations in a sensible health topic would be ethically questionable.

Table A.1: Importance of reasons for Uncertainty Resolution

Group	Factual knowledge	Social Acceptance	Other	observations
(1) All participants	3.84 (270)	3.19 (257)	2.72 (50)	300
(2) Self-proclaimed	4.13 (169)	3.49 (159)	2.84 (25)	228
(3) Strong self-procl.	4.52 (84)	3.87 (77)	2.71 (14)	86
(4) Stated at 50€	4.01 (170)	3.39 (160)	3.18 (33)	184
(5) Stated at 800€	4.41 (54)	3.86 (51)	3.56 (9)	57
(6) Stated & self-procl. at 50€	4.24 (114)	3.60 (109)	3.35 (17)	120
(7) Stated & self-procl. at 800€	4.41 (41)	3.95 (38)	3.71 (7)	43
(8) Stated & strong self-procl. at 50 €.	4.54 (68)	3.94 (62)	3 (12)	70
(9) Stated & strong self-procl. at 800€	4.59 (32)	4.1 (30)	3.5 (6)	33

A.3 Robust results

P-MLM regression results for the robust sample

Table A.2: Results of Mixed Logit Choice Models for robust sample

Predictor	Model 6	Model 7
No diagnosis	(base alternative)	(base alternative)
Diagnosis		
Co-pay	-0.0059***	-0.0062***
	(-0.0074, -0.0043)	(-0.0078, -0.0046)
Treatment accuracy	0.0276*	0.0312**
	(0.0044, 0.0509)	(0.0074, 0.0550)
Normal		
SD Co-pay	0.0039	0.0043
	(0.0026, 0.0057)	(0.0030, 0.0062)
Corr. Co-pay x tr. acc.	0.7891	0.6321*
	(-0.8051, 0.9970)	(-0.1576, 0.9287)
SD Treatment acc.	0.0179	0.0208
	(0.0095, 0.0336)	(0.0128, 0.0337)
Diagnostic test		
Symptom level	-0.4219	-0.2882
	(-1.6969, 0.8532)	(-1.6115, 1.0351)
Age	0.7071	
	(-0.6859, 2.1001)	
Public insurance	-1.3439	
	(-3.7979, 1.1101)	
Higher education	0.8388	
	(-0.8101, 2.4876)	
Household Income	-0.3679	-0.4298
	(-1.0672, 0.3313)	(-1.0954, 0.2359)
EMS	1.4722	1.5601
	(-0.3738, 3.3183)	(-0.3706, 3.4908)
Anxiety	-0.0261	0.0367
	(-0.6565, 0.6044)	(-0.6370, 0.7104)
Actual symptoms	-0.3184***	-0.3999***
	(-0.5283, -0.1086)	(-0.6257, -0.1742)
Low perceived need	0.2961	0.3238
	(-0.3755, 0.9677)	(-0.3957, 1.0433)
Discomfort	0.3687	0.3947

Note. 95% confidence intervals are recorded in parentheses. Significance levels: * = .05, ** = .01, *** = .001.

Table A.2 – *Continued from previous page*

Predictor	Model 6	Model 7
	(-0.2405, 0.9780)	(-0.2327, 1.0220)
Self-reported UR	0.6332**	0.6448*
	(0.1442, 1.1222)	(0.1397, 1.1499)
Language	0.2053	
	(-2.0632, 2.4737)	
Constant	1.7997	2.6437
	(-3.8950, 7.4943)	(-1.6036, 6.8910)
Specialist Doctor		
Symptom level	-0.0852	-0.0028
	(-1.2735, 1.1030)	(-1.2372, 1.2316)
Age	0.5145	
	(-0.7794, 1.8084)	
Public insurance	-1.0171	
	(-3.3830, 1.3488)	
Higher education	0.6451	
	(-0.8588, 2.1489)	
Household Income	-0.1656	-0.1933
	(-0.8107, 0.4796)	(-0.8103, 0.4238)
EMS	1.8193*	1.9147*
	(0.0892, 3.5494)	(0.1005, 3.7289)
Anxiety	0.0604	0.1056
	(-0.5252, 0.6461)	(-0.5207, 0.7318)
Actual symptoms	-0.3514***	-0.4289***
	(-0.5495, -0.1533)	(-0.6440, -0.2137)
Low perceived need	0.2496	0.2971
	(-0.3743, 0.8735)	(-0.3744, 0.9686)
Discomfort	0.3136	0.3712
	(-0.2580, 0.8851)	(-0.2181, 0.9606)
Self-reported UR	0.0090	0.0189
	(-0.4419, 0.4600)	(-0.4474, 0.4851)
Language	0.7971	
	(-1.3795, 2.9736)	
Constant	1.8334	2.3545
	(-3.4828, 7.1497)	(-1.6316, 6.3406)
Log-likelihood	-634.0009	-638.6348
AIC	1330.002	1323.27
BIC	1496.3229	1446.669

Note. 95% confidence intervals are recorded in parentheses. Significance levels: * = .05, ** = .01, *** = .001.

Robust results for hypothesis 4

Table A.3: Respective share of patients choosing no diagnosis instead of the stated alternative

Alternative	Relative (%)	Absolute
Specialist at 80% treatment accuracy		
All	3.16	5
Specialist at 100% treatment accuracy		
All	1.27	2
Diagnostic test at 200€ co-pay		
All	13.92	22
Diagnostic test at 800€ co-pay		
All	37.34	59
Sample size		158

Appendix B

Programming code

The variables are named as follows, listed in the alphabetic order:

Table B.1: Variable description for the Stata code

Variable name	Description
age_num	Indicates the age category in numeric order
anxiety	The self-reported anxiety level
att1_num	The co-pay in euros
att2_num	The treatment accuracy in percent
attitude_1	The score of having a low perceived need for doctors visits
attitude_2	The score of having discomfort for doctors visits
attitude_3	The score of a self-reported preference for UR
ct_num	The choice task number, hence the panel identifier in the DCE
diagnosis	Stands for the diagnostic pathway (categorical)
Edu_cat	Education category as inquired in the survey
ems	Self-report of having received an EMS diagnosis by a professional, equal to 1 if yes is indicated
hhi_num	Indicates the approximate household income in euros
ID	Identifier of subject participating in the study
insurance	Is a categorical variable accounting for the type of insurance, PHI, SHI, or outside of Germany
language	Is a dummy variable that is equal to 0 for answering the survey in German and 1 for answering it in English
Level	A categorical variable for the hypothetical symptom level
selected_choice	The dummy variable for a choice being selected, hence the dependent variable in all regressions
sym	The sum of all symptoms

Code:

```
/**** Master thesis: Uncertainty Resolution in EMS Diagnosis
Author: Stephan Rothenberger

****/
```

clear all

import final survey data that was already filtered for valid responses and formatted partly in Excel
import delimited "ShapedSurveyData.csv" , delim(",") varnames(1) stripquote(yes) case(preserve) encoding(UTF-8)
*Create a numeric ID per participant

egen ID = group(id)

drop id

*understand missing values
misstable summarize

*Add missing values to the lowest category for education and to not diagnosed for EMS. Further, all missing/prefer not to say symptom and attitude values are categorized to contradict any hypothesis, hence, if at all, weakening the result, to ensure we do not manipulate towards a stronger effect.

mvencode Edu_cat, mv(0) override
mvencode ems, mv(0) override
mvencode anxiety, mv(0) override
mvencode attitude_1, mv(3) override
mvencode attitude_2, mv(1) override
mvencode attitude_3, mv(1) override

*encoding all string variables and creating a meaningful predictor of EMS risk

encode Insurance, gen(insurance)
drop Insurance

encode UserLanguage, generate(language)
drop UserLanguage

encode hhi, gen(hhi_num)
encode att5, gen(diagnosis)
drop att5

*group all older than 45 y.o. into one group due to small prevalence and high chance of menopause (thus not suffering from ems)

recode age_num 5 = 4
recode age_num 6 = 4

*generate Variable for Symptoms, thereby only taking into account women who have a minimum symptom level of 2 (moderate) in their subjective pain assessment or a minimum impact on daily life assessment of 2 (moderate)

generate sym = Symp_sum*Symp_dummy
tab sym

*set ID as the identifier for panel data units and choice task (ct_num) for ”time periods” and alternative identify (diagnosis type)

cmset ID ct_num diagnosis

********** Analysis **********

**********Descriptive Statistics**********

xtsum

tab age_num
tab insurance
tab Edu_cat
tab hhi_num
tab language

* basic MNL regression, note that this model does not include the description/type of diagnosis as it has collinearity with the price and accuracy of treatment when including all trade-offs

*model including all exogenously varied variables
xtmlogit selected_choice att1_num att2_num i.Level,
estimate store BaseMNLSlim
estat ic

*model adding all attitude variables

xtmlogit selected_choice att1_num att2_num i.Level i.insurance age_num i.Edu_cat hhi_num i.language i.ems sym anxiety attitude_1 attitude_2 attitude_3
estimate store BaseMNLExtended
estat ic

*-¿ no benefit, as expected as explanatory variables are not allocated to trade-offs or alternatives, hence these cannot be accounted for properly

***mixedlogit
*note that this model does not include the description/type of diagnosis as it has collinearity with the price and accuracy of treatment when including all trade-offs

*model including all exogenously varied variables
cmxtmixlogit selected_choice, random(att2_num, correlated) casevars(i.Level) basealternative (No diagnosis) random(att1_num, correlated) intpoints(150)

estimate store BasePMLMSlim
estat ic

*Extended Model adding all attitude variables
cmxtmixlogit selected_choice, casevars(i.ems sym i.Level i.insurance hhi_num Edu_cat age_num i.language anxiety attitude_1 attitude_2 attitude_3) random(att1_num, correlated) random(att2_num, correlated) basealternative (No diagnosis) intpoints(150)

estimate store BasePMLMExt1
estat ic

***Only attitude and health and hhi
cmxtmixlogit selected_choice, casevars(i.ems sym i.Level hhi_cat anxiety attitude_1 attitude_2 attitude_3) random(att1_num, correlated) random(att2_num, correlated) basealternative (No diagnosis) intpoints(150)

estimate store BasePMLMImproved
estat ic

****
*overview

margins, dydx(*)

**********Hypothesis 1**********

*A majority of individuals make the cost-minimizing, benefit-maximizing choice of going to an endometriosis specialist if there is no improvement in treatment outcome.

*Uncertainty resolution at 50€
tabulate alt if ct_num == 9 & selected_choice == 1
tabulate diagnosis if ct_num == 9 & selected_choice == 1
*Uncertainty resolution at 800€
tabulate alt if ct_num == 10 & selected_choice == 1
tabulate diagnosis if ct_num == 10 & selected_choice == 1

*t-test to verify the difference of means

ttest alt==1.5 if ct_num == 9 & selected_choice == 1
ttest alt==1.5 if ct_num == 10 & selected_choice == 1

********** Hypothesis 2 *********
*The diagnostic test is chosen more often with decreasing price.

*Hyp2
*interpret general model
estimate restore BasePMLMImproved

*Interpret stated preferences at 80%. Choice tasks: 5 (200€) 6 (800€)
tab diagnosis if selected_choice == 1 & ct_num==5
tab diagnosis if selected_choice == 1 & ct_num==6

*Margins of the price at population average accuracy

margins, alternative("Diagnostic test - Secured diagnosis") at(att1_num=(800 400 200 100 50))

**********Hypothesis 3 & 4**********

*Increasing symptoms do not impact the choice of diagnostic method, and therefore the uptake in diagnostic tests.
*The ostrich effect, which refers to the tendency to avoid negative information, is not a significant
factor in the decision between diagnostic options for EMS, contrary to most indications.

OSTRICH EFFECT & INFORMANTION AVOIDANCE

**Descriptive statistics for increasing symptoms (H4) and ostrich effect (H5)
*avoidance of specialist doctor at no cost

*at 80% probability of correct treatment

tabulate diagnosis if ct_num == 1 & selected_choice == 1

tabulate diagnosis if ct_num == 1 & Level == 1 & selected_choice == 1
tabulate diagnosis if ct_num == 1 & Level == 2 & selected_choice == 1

*at 100% probability of correct treatment

tabulate diagnosis if ct_num == 2 & selected_choice == 1

tabulate diagnosis if ct_num == 2 & Level == 1 & selected_choice == 1
tabulate diagnosis if ct_num == 2 & Level == 2 & selected_choice == 1

*avoidance of diagnostic test at cost
*200€ cost

tabulate diagnosis if ct_num == 3 & selected_choice == 1

tabulate diagnosis if ct_num == 3 & Level == 1 & selected_choice == 1
tabulate diagnosis if ct_num == 3 & Level == 2 & selected_choice == 1

*800€ cost tabulate diagnosis if ct_num == 4 & selected_choice == 1

tabulate diagnosis if ct_num == 4 & Level == 1 & selected_choice == 1
tabulate diagnosis if ct_num == 4 & Level == 2 & selected_choice == 1

*Willingness to pay with symptoms
estimates restore BasePMLMImproved

margins, dydx(Level)

**********Hypothesis 5**********
*If subjects do choose not to get diagnosed, their choice is highly correlated with factors of information avoidance in terms of, anxiety, discomfort, or low perceived need of going to the doctor, i.e., avoiding a diagnosis for one or multiple of these reasons.

*Drivers of information avoidance

*information avoidance

estimate restore BasePMLMImproved

margins if ct_num < 5, dydx(anxiety)
margins if ct_num < 5, dydx(attitude_1)
margins if ct_num < 5, dydx(attitude_2)

UNCERTAINTY RESOLUTION
*Descriptive statistics of Uncertainty Resolution

**********Hypothesis 6**********
*Some individuals seek information on their health, even though there is no economic benefit to it.

**Stated preference
*Uncertainty resolution at 50€
tabulate diagnosis if ct_num == 9 & selected_choice == 1

*Uncertainty resolution at 800€
tabulate diagnosis if ct_num == 10 & selected_choice == 1

**Stated preference & self-proclaimed preference for UR
*Uncertainty resolution at 50€
tabulate diagnosis if ct_num == 9 & attitude_3 > 1 & selected_choice == 1

*Uncertainty resolution at 800€
tabulate diagnosis if ct_num == 10 & attitude_3 > 1 & selected_choice == 1

**Stated preference & self-proclaimed preference for UR
*Uncertainty resolution at 50€
tabulate diagnosis if ct_num == 9 & attitude_3 > 3 & selected_choice == 1

*Uncertainty resolution at 800€
tabulate diagnosis if ct_num == 10 & attitude_3 > 3 & selected_choice == 1

margins, dydx(attitude_3)
margins, at(attitude_3=(1 5))

**********Hypothesis 7**********

* With increasing symptoms, patients' preference to resolve uncertainty does not increase.

*Descriptive stated preference

*Uncertainty resolution at 50€
tabulate diagnosis if ct_num == 9 & Level == 1 & selected_choice == 1
tabulate diagnosis if ct_num == 9 & Level == 2 & selected_choice == 1

*Uncertainty resolution at 800€
tabulate diagnosis if ct_num == 10 & Level == 1 & selected_choice == 1
tabulate diagnosis if ct_num == 10 & Level == 2 & selected_choice == 1

*Margins
estimate restore BasePMLMImproved

margins if ct_num > 8, dydx(Level)

**********Hypothesis 8**********
*Social validation is the key driver for uncertainty resolution.

*entire population
mvtest means reasons_1 reasons_2 reasons_3
*individuals with any self-proclaimed preference for UR
mvtest means reasons_1 reasons_2 reasons_3 if attitude_3 > 1
*individuals with strong self-proclaimed preference for UR
mvtest means reasons_1 reasons_2 reasons_3 if attitude_3 > 3

*individuals with strong self-proclaimed preference for UR and stated preference at 50€
mvtest means reasons_1 reasons_2 if attitude_3 > 3 & ct_num == 9 & selected_choice == 1 & alt == 2
*individuals with strong self-proclaimed preference for UR and stated preference at 800€
mvtest means reasons_1 reasons_2 if attitude_3 > 3 & ct_num == 10 & selected_choice == 1 & alt == 2

*Model 1
xtsum reasons_*

*Model 2
*any self-proclaimed preference for UR
xtsum reasons_* if attitude_3 > 1

*Model 3

*Strong self-proclaimed preference for UR
xtsum reasons_* if attitude_3 > 3

*Model 4
*Stated preference for UR
xtsum reasons_* if ct_num == 9 & selected_choice == 1 & alt == 2

*Model 5
*Strong stated preference for UR
xtsum reasons_* if ct_num == 10 & selected_choice == 1 has& alt == 2

*Model 6
*Stated and self-proclaimed preference for UR
xtsum reasons_* if attitude_3 > 1 & ct_num == 9 & selected_choice == 1 & alt == 2

*Model 7
*Strong stated and self-proclaimed preference for UR
xtsum reasons_* if attitude_3 > 1 & ct_num == 10 & selected_choice == 1 & alt == 2

*Model 8
*Stated and strong self-proclaimed preference for UR
xtsum reasons_* if attitude_3 > 3 & ct_num == 9 & selected_choice == 1 & alt == 2

*Model 9
*Strong stated and strong self-proclaimed preference for UR
xtsum reasons_* if attitude_3 > 3 & ct_num == 10 & selected_choice == 1 & alt == 2

********** Robustness check **********

*everything repeated with a core sample of 158 German females under 35, dropping samples that did not indicate education, hhi, EMS, symptoms, and attitude values.

clear all

*import final survey data that was already filtered for valid responses and formated partly in excel

import delimited ”ShapedSurveyData.csv” , delim(”,”) varnames(1) stripquote(yes) case(preserve) encoding(UTF-8)

*create numeric ID per participant

egen ID = group(id)

drop id

*understand missing values
misstable summarize

*Filter values so only the core sample for robustness check stays

drop if missing(Edu_cat)
drop if missing(ems)
drop if missing(hhi_num)
drop if missing(anxiety)
drop if missing(Symp_1)
drop if missing(Symp_2)
drop if missing(attitude_1)
drop if missing(attitude_2)
drop if missing(attitude_3)
keep if age_num<3
drop if strpos(Insurance, ”No, I’m insured in”) > 0

*encoding all string variables and creating a meaningful predictor of EMS risk

encode Insurance, gen(insurance)
drop Insurance

encode UserLanguage, generate(language)
drop UserLanguage

encode hhi, gen(hhi_num)
encode att5, gen(diagnosis)
drop att5

*generate Variable for Symptoms, thereby only taking into account women who have a minimum symptom level of 2 (moderate) in their subjective pain assessment or a minimum impact on daily life assessment of 2 (moderate)

generate sym = Symp_sum*Symp_dummy
tab sym

*set ID as the identifier for panel data units and choice task (ct_num) for ”time periods” and alternative identify (diagnosis type) cmset ID ct_num diagnosis

********** Analysis **********

The analysis is the same as above.