DEVELOPMENT OF A SCALE FOR

EVALUATING SCREENING TESTS FOR CANCER:

ATTRIBUTES PATIENTS EMPHASIZE

(ESCAPE)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A Dissertation Proposal

presented to

the Dissertation Committee

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

by

K.D. VALENTINE

Dr. Victoria Shaffer, Dissertation Supervisor

MAY 2019

The undersigned, appointed by the dean of the Graduate School have examined the dissertation entitled

DEVELOPMENT OF A SCALE FOR

EVALUATING SCREENING TESTS FOR CANCER:

ATTRIBUTES PATIENTS EMPHASIZE

(ESCAPE)

presented by KD Valentine

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Associate Professor Victoria Shaffer

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Professor Phillip Wood

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Professor Denis McCarthy

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Doctor Richelle Koopman

DEDICATION

First, I would like to thank my parents Richard Valentine and Debbie Valentine for being supportive (both emotionally and financially) over these many years. Without their belief that I could do whatever I set my mind to, and their push to make my dreams of earning my PhD a reality, I would not be here today. Second, I would like to thank my faithful companions, Archibald Blankie Valentine and Warlock Valentine for being by my side through these past few years. Their kind eyes, reassuring snugs, and happy, wagging tails made this process infinitely easier. Finally, I would like to thank my family, friends, and peers who have helped me along the way. From phone calls to late nights, you all have made this all possible. Thank you.

ACKNOWLEDGEMENTS

I would like to thank the many professors who have helped pave my way to this degree. First, I would like to thank Dr. Harry L. Hom for allowing me to be a member of his research team in my first year at Missouri State University, for challenging me to write and think beyond my comfort zone at the time, and for ensuring I entered into a graduate program under the mentorship of Dr. Erin M. Buchanan. Second, I would like to thank Dr. Erin M. Buchanan for taking me on as her first graduate student, for mentoring me during and after my Master’s program, for putting up with my loud yawns, and for ensuring I was prepared for the world of academia. You taught me to learn, teach, and make the most of my research potential. I would like to thank Dr. Laura D. Scherer for allowing me to complete my research rotation in her lab, for allowing me to be a part of her excitement and inquiries, and for helping me get to where I am today. I would like to thank Dr. Victoria A. Shaffer for accepting me as her graduate student, for advising me these past four and a half years, and for making me such a success. Finally, I would like to thank Dr. Mark Hannink and Debbie Allen for my Life Sciences fellowship that has allowed me to pursue a research line that interested me.

TABLE OF CONTENTS

Acknowledgements iv

Chapter

1. Chapter one 1

Biases

Accuracy

Changing Guidelines

Overenthusiasm

Preference Sensitive Screening

Understanding Preferences for Screening

Qualitative Measures of Preference

Quantitative Measure of Preferences

Individual Differences Measures

Need for a New Individual Difference Measure

1. Chapter two 22

Item Development

1. Chapter three 26

Exploratory Factor Analysis

Rationale

Participants and Methods

Analysis

Results

Conclusions and Limitations

1. Chapter four 32

Confirmatory Factor Analysis

Rationale

Participants and Methods

Analyses

Results

Conclusions and Limitations

1. Chapter five 37

Assessing the Reliability of the ESCAPE Questionnaire

Rationale

Participants and Methods

Analyses

Results

Conclusions and Limitations

1. chapter six 40

Assessing the Validity of the ESCAPE Questionnaire

Rationale

Participants and Methods

Analyses

Results

Conclusions and Limitations

1. Chapter seven 45

Scale Generalization Study

Rationale

Participants and Methods

Analysis

Results

Conclusions and Limitations

1. Chapter eight 52

Comparison of Stated Preference Methods Study

Rationale

Participants and Methods

Analysis

Results

DCE v. TT Threshold Values

Relationships between DCE, TT, and ESCAPE

Predicting Threshold and ESCAPE Values

Predicting Choice with Thresholds and ESCAPE

Conclusions and Limitations

1. Chapter nine 64

Discussion and Future Directions

REFERENCES 70

Appendix

1. LITERATURE SEARCH PROCESS 84
2. Demographics for all studies 86
3. original list of items by theoretical category 87
4. descriptive statistics for disease EFa 90
5. ALTERNATIVE ROTATIONS FOR UNDERGRADUATES 94
6. Descriptive statistics for efa 97
7. scree plot for undergraduate efa 99
8. Apriori hypotheses 100
9. ALTERNATIVE ROTATIONS FOR MTURKERS 103
10. parallel analysis AND scree plots 106
11. low quality data identification and handling 107
12. descriptive statistics for efa mturk data 109
13. calculating dce and tt thresholds 112
14. undergraduate low quality data detection 113

Vita 114

DEVELOPMENT OF A SCALE FOR

EVALUATING SCREENING TESTS FOR CANCER:

ATTRIBUTES PATIENTS EMPHASIZE

(ESCAPE)

K.D. Valentine

Dr. Victoria Shaffer, Dissertation Supervisor

ABSTRACT

The public’s overenthusiasm for cancer screening tests has the potential to subject many individuals to harms such as overdetection and false positives. Thus, it is of the utmost importance that we understand what drives patient preferences for screening in the first place. Then, once these preferences are defined and understood, the information regarding the various positive and negative attributes of options—as well as the likelihoods associated with these attributes for an individual—and any other features of the decision could be tailored to the individual. This dissertation proposes and validates a new measure that identifies what features individuals find important when choosing a screening test and how they vary relative to others. Herein, a set of factors regarding screening test attributes was created, and a 5-factor structure was both explored and confirmed. The scale is shown to be reliable and to have convergent and discriminant validity. Further, the structure was not found to replicate in a more diverse population. Instead the more diverse sample has a 6-factor structure. Finally, this individual difference scale was compared with a discrete choice experiment and a threshold technique, finding all of these methods vary and none of them are capable of predicting screening choices.

**Chapter 1**

Screening tests are measures of secondary prevention that aim to detect cancers in the early stages, with the assumption that this will either extend life or increase the chance of survival. Screening technologies have made great advances in recent years, contributing to the increase in cancer survival rates from 50% to 64% between 1976 and 2006 (Herbst et al., 2006). However, the decision to screen is a multifaceted one. A patient must process and apply information about (a) lead and length-time biases, (b) test accuracy and its implications for overdiagnosis and overtreatment, (c) changing guidelines that often recommend reducing or stopping screening, (d) societal overenthusiasm for screening, and (e) the changing nature of the physician-patient relationship, moving from paternalism to shared decision making.

**Biases**

Lead-time and length-time biases influence the evidence of screening test efficacy. Lead time bias occurs when a screening test simply detects a cancer earlier, but does not actually extend the person’s life (Feinleib & Zelen, 1969). For example, if an individual did not screen for cancer, the cancer was found at age 50, and the individual died at age 53, we would say this person survived 3 years. However, if the same person were screened, that cancer could be found earlier, say at age 45, and they still die at 53. Now that same individual’s survival is 8 years. Ideally, an effective screening test should extend life. When evaluating a screening test with an outcome of survival time, these estimates are subject to lead-time bias that can favor a screening test when there is not actual benefit to screening.

Length-time bias occurs because of the rate of growth of cancers (Esserman, Thompson, & Reid, 2013). Cancers grow and spread at different speeds. So, for example, imagine one person has a very slow growing cancer and another has a very aggressive, fast growing cancer. Given the rapid growth, the second person dies from cancer, whereas the person with the slow-growing cancer lives to the point of being screened. There is a shorter period of time during which the second person has cancer which could be detected by a screening test but the cancer is not yet large enough to be symptomatic or cause the person to be diagnosed without screening. Because of this, screening tests are more likely to detect slow growing cancers. If slow growing cancers cause fewer deaths than fast growing cancers, people who have cancer detected through screening will do better (on average) than those whose cancer was detected through symptoms—even if there is no actual benefit in catching the cancer earlier through screening. This can lead to the impression that screening makes cancer less dangerous when in fact it is actually just less dangerous cancers that are being detected by screening. Taken together, these biases can arbitrarily inflate the perceived benefit of screening tests. For example, Shwartz (1980)compared the five-year survival rates of women whose breast cancer was found via screening (80% survived) to those women whose breast cancer was diagnosed clinically (50% survived). They suggest that about 50% of the increase in five-year survival rates for those whose cancer was detected through screening may be due to these biases.

**Accuracy**

Further complicating patients’ decisions about screening is information about test accuracy. Developers of each screening test must set a threshold for diagnosis that provides an acceptable balance between the rate of false positives and the rate of false negatives. If the goal of the test is to make sure that everyone with cancer is told they have cancer, this means that the test will also probably tell quite a few people without cancer that they do have cancer (Swets & Swets, 1988). This is called a false positive. This false positive is due to the sensitivity of the test. If a test is highly sensitive it will ensure that none with cancer are misdiagnosed as not having cancer, but it will also state that some people have cancer when in reality they do not. If instead, the goal of the screening test is to minimize these false positives, this necessarily increases the risk of false negatives (Swets & Swets, 1988). A false negative means that the test incorrectly tells a person with cancer that they do not have cancer. This is due to the specificity of the test. If the test is highly specific, it will ensure that no one without cancer is falsely diagnosed as having cancer, however it will also tell some people that they do not have cancer, when in fact they actually do have cancer. It is important to note that when the results of a screening test are presented to a patient we do not know what is true in real life (this means that we do not know if a person has been correctly or incorrectly diagnosed until later, after more tests and time are involved). This necessarily increases the complexity of patient risk communication.

What an individual usually wants to know when they receive their results is, given a screening test has told me that I have cancer, what is the probability that I actually have cancer? This is called the positive predictive value (PPV) of a test, and its calculation requires information about the prevalence of the cancer, the specificity of the test, and the sensitivity of the test. PPV rates are usually much smaller than the sensitivity of the test, but many individuals confuse these two values (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007; Paulos, 1988). These values are quite variable (Rosenberg et al., 2006) and are not usually presented to patients.

An additional concern arises from the fact that screening tests can detect three progressions of cancers (Kattan, 2016). A test may detect 1) a cancer that is so swift growing that no actions can alter the prognosis, 2) a cancer that grows at a moderate rate that can grow to be symptomatic and life threatening, but which can be remediated if detected early enough, or 3) a cancer that is so slow growing that it would never develop into anything symptomatic or life-threatening. The latter case is called overdetection (detecting a cancer that would otherwise never have been noticed or concerning), which can lead to overtreatment (receiving treatment for a condition that is unnecessary). While some cancers are known to be swift growing (e.g. pancreatic cancer) or slow growing (e.g. thyroid cancer), others may progress in less predictable ways. However, these facts are not always made clear to the public; thus, people may believe that regardless of the progression, action must be taken to treat the cancer.

Additionally, we must realize that this is not a simple yes or no decision that an individual makes once. People have to choose whether or not to begin screening, at what age to begin screening, how frequently they will be screened, which test they will use to screen (when there are multiple options), and when to stop screening. Therefore, the tradeoff between the risk of a false positive or false negative diagnosis—or even overdetection and overtreatment—and the potential benefit in terms of increased life expectancy, must be weighed to determine if a screening tests can be considered appropriate for a given individual at a given time.

**Changing Guidelines**

Guidelines and recommendations are designed to help patients balance information about the benefits and risks of a specific test. Groups such as the United States Preventive Services Task Force (USPSTF) are tasked with synthesizing research on the effectiveness of various screening tests. These groups use this information to create a set of guidelines designed to maximize public health impact. These panels consider personal risk (including family history and lifestyle choices) and age to specify a recommendation.

However, guidelines are not static, and the continuous updating of these recommendations adds an additional layer of complexity to these decisions. These panels routinely consider the most recent research and regularly update the recommendations. For instance, screening tests, such as chest x-rays to screen for lung cancer were previously believed to effectively prevent death from cancer, but are now understood to have definitive harms for nonsmokers (including high rates of overdiagnosis) and tentative benefits associated with them for smokers (Welch & Black, 2010). For other tests, such as the prostate specific antigen (PSA) test, the USPSTF previously did not recommend that any men receive this test (Moyer, 2012); however, these guidelines underwent revisions (Bibbins-Domingo, Grossman, & Curry, 2017) and now suggest that men 55-69 years participate in shared decision making with their physician regarding being screened in order to weigh their individual risks and benefits (Grossman et al., 2018).

These recommendations are only useful when followed; however, physicians and patients have both been slow—or even unwilling—to adopt new recommendations that focus on reducing or stopping screening. For example, in 2009, the USPSTF released new guidelines regarding the use of mammography to screen for breast cancer. The guideline, which previously recommended mammography every 1-2 years, no longer recommended routine screening for women aged 40-49 years (DeAngelis & Fontanarosa, 2010). This revision of recommendations led to what some have deemed a “firestorm” around breast cancer screening (Laine, Dickersin, & Mulrow, 2016). The public was exposed to the change in recommendations through newspapers (Ehrenreich, 2009; Kamerow & Woolf, 2009) and editorials (DeAngelis & Fontanarosa, 2010; Woloshin & Schwartz, 2010), and the controversy over the changes and their potential effects on healthcare reform were paraded on news and social media. A content analysis by Squires and colleagues (2011) showed that over half (51.9%) of posts on social media and reports on the news did not support the recommendation; only 17.6% were supportive of the recommendation. Further investigation by Squires and colleagues (2011) showed that this disagreement ultimately led to confusion. When conducting a survey of women aged 40-74 with no history of breast cancer they found that, 30% of women reported feeling confused by the recommendation. This confusion was most pronounced among the women between the ages of 40-49—the exact group the change in recommendation was targeted toward. It is therefore unsurprising that a study done by Howard and Adams (2012) analyzing the Medical Expenditure Panel Surveys from 2006-2010 showed no difference in rates of mammography when comparing the years before vs. after the recommendation. Clearly this recommendation did not change behavior.

Similarly, in 2009 when two large clinical trials were released which reported that routine PSA testing resulted in surprisingly high overdiagnosis rates many organizations reacted by changing their recommendations for PSA screening. Following this, Goodwin and colleagues (2013) investigated a commercial insurance database for PSA testing claims filed from 2001 to 2011. They analyzed records to determine whether PSA screening had actually decreased since the release of the new recommendation. They found that the slope of rates of PSA testing did not change for men age 50-64 from before to after the recommendation change. Further, while men 40-49 showed a slow increase in slope up to 2008, this slope became flat from 2009 to 2011, but never decreased. Overall, these results indicated a lack of uptake of the new guidelines. When the USPSTF further altered their recommendations in 2012 to recommend against PSA testing for all men, this failure to update could again be seen. In a study by Perez and colleagues (2015) the number of individuals referred to urologist in their department for elevated PSA levels was investigated for the 12 month period before and after the USPSTF guidelines were reported. They, too, found no decline in PSA testing that would indicate that the new guidelines were being followed.

It is possible that continued usage of PSA may be due to physician recommendations and that patients may prefer not to be screened; previous work has demonstrated that poor consistency between patient preferences and screening test choices can be at least partially attributed to physician recommendations (Schroy et al., 2011, p. 186). However, it is also possible that the patient themselves want the screening test, regardless of the recommendation. This latter explanation has been supported by work by Squires and colleagues (2013) which presented men (without a history of prostate cancer) with the new USPSTF’s draft recommendation on PSA—the USPSTF recommended against PSA testing for men of all ages. While the majority of men (62%) stated that they agreed with the recommendation, only 13% of men stated that they actually planned to follow the recommendation. The authors attribute this result, in part, to the fact that the media routinely touts cancer screening as a way to save lives; thus, the public is said to already be “sold on the benefits of screening” (Squires et al., 2013).

**Overenthusiasm**

Unsurprisingly, the US population has repeatedly been shown to be quite enthusiastic when it comes to cancer screening (Schwartz, Woloshin, & Fowler Jr, 2004; Waller, Osborne, & Wardle, 2015). Work by Schwartz and colleagues (2004) demonstrated that in individuals with no history of cancer, the majority approve of routine screening (87%) and many (74%) believe that if cancer is detected early it saves that person’s life most or all of the time. Further, although 38% stated they had experienced a false positive test, and more than 40% of those classified that experience as a frightful experience (i.e. "very scary" or the "scariest time of my life"; Schwartz et al., 2004, p. 75), almost all (98%) stated they “were glad they had the initial screening test” (Schwartz et al., 2004, p. 75). Here, even after experiencing the negative consequences associated with a false positive, they do not seem to regret their decision.

This overenthusiasm has previously been tied to the population being under informed about the risks and benefits screening tests carry (Hoffman et al., 2016). This can be seen in studies like one by Shokar and colleagues (2005) which showed that participant’s knowledge of cancer (i.e. types, warning signs or symptoms, risk factors) and cancer screening (i.e. describing what screening means or defining it) was quite poor. They report that, “none of the subjects realized that the purpose of screening is to detect disease in the absence of symptoms” (Shokar et al., 2005, p. 344), but these same individuals reported believing in the screening benefit of early diagnosis. This work is supported by previous work by Steele and colleagues (2000) who surveyed men and found that 1/3 of men did not know what their risk of prostate cancer was, and further, that almost half of their sample had never heard of PSA testing. Work regarding breast cancer by Davis and colleagues (1996) revealed that 22% of their convenience sample of women visiting outpatient clinics did not know why women should receive mammograms, 72% reported not knowing when women should begin having mammograms, and 75% did not know the frequency with which a mammogram should be conducted. More recent work by Dolan and colleagues (2004) sampled veterans aged 50 or older and found that about 47% of participants had not heard of tests that find colorectal cancer. Yet, over 92% of these participants stated that they believed their chances of survival were “good to excellent” if the cancer was found early (Dolan et al., 2004, p. 2620).

However, even when individuals are well informed about the risks and benefits associated with screening tests, they may still choose to be screened in situations where the probability of harms are quite high while the benefits are quite low (Dolan & Frisina, 2002; Dolan et al., 2004; O’Connor et al., 1999). A study by Wilt and colleagues (2001) mailed a leaflet to male veterans recruited from a primary care clinic with information about PSA testing, including information recommending against PSA testing for these older men stating, “Many doctors think older men and men with serious medical problems should not have a PSA” (Wilt et al., 2001). The men who received the pamphlet were better able to correctly state that most men with early prostate cancer who do not take action will not die, that prostate cancer treatment does not necessarily lengthen life, and that most men with a high PSA level do not have prostate cancer as compared to a group who received no pamphlet—just usual care. Despite the fact that the intervention improved knowledge, it did not seem to change the fact that equal proportions of these men received PSA testing within the next year (31-37%), regardless of whether they received the pamphlet or usual care. This seems to reflect a tendency of these individuals to desire the test even when warned of the disproportionate amount of risk as compared to benefits, perhaps because they personally evaluate any chance of benefit as logically being worth the related risks. This may be due, in part, to motivated reasoning. When a person is confronted with information that is discordant with their preexisting beliefs, they may ignore this discordant information or rationalize why their previously held beliefs are still correct by focusing on the more positive information. Here, this means that an individual may well be informed of the risks of the test, but may instead choose to focus on the possible benefits and ignore the new risk information, ultimately deciding that screening is worth it.

One study, which took this enthusiasm to be screened to a logical extreme, asked various samples of participants if they would be interested in receiving a hypothetical screening test when research had, “unquestionably shown that the test does not extend life or reduce the chance of death from…cancer” (Scherer, Valentine, Patel, Baker, & Fagerlin, 2018, p. 12). Overall, about half of participants stated that they would choose to take the screening test even in the absence of benefit. Interestingly, this study showed that many individuals stated they did not believe that the screening test actually had no benefits, which could either reflect that participants did not trust the information, or that participants were already “sold on the benefits of screening” (Squires et al., 2013, p. 186). But, when considering those that did believe the test had no benefits, still about a third stated they wanted to be screened by the unbeneficial screening test. Participants who feel they are at high risk of cancer, are worried about cancer, or believed the test reduced cancer deaths were more likely to indicate a willingness to be screened by this unbeneficial test. In contrast, those who thought the test was risky were less likely to choose the test. Investigation into qualitative responses by participants indicated that even in the absence of the benefit of reduction in cancer death, the idea that the test would provide them with some sort of health information made the test seem worthwhile.

This potentially dangerous overenthusiasm for some tests can lead to individuals being screened for cancer when they are unlikely to benefit from the test but are still at risk of a false positive diagnosis or overdiagnosis. These can cause emotional harms in the form of anxiety (Scaf-Klomp, Sanderman, van de Wiel, Otter, & van den Heuvel, 1997); physical harms including pain, hospitalization, and even death (Mulhem, Fulbright, & Duncan, 2015); and financial harms in the form of debt accrued due to medical procedures and missed work (Victory, 2017).

**Preference Sensitive Screening**

Given that the public seems to be “sold on the benefits of screening” (Squires et al., 2013, 186), this would suggest that without any intervention or tailoring of information that decisions about screening might largely default to pro-screening decisions. This bias toward action is particularly important to understand as guidelines are increasingly recommending preference sensitive screening (Oeffinger et al., 2015; U.S. Preventive Services Task Force, 2016; Wolf et al., 2010). The American Cancer Society’s Cancer Screening Guidelines webpage now suggest that individuals, “talk with a health care provider about the uncertainties, risks, and potential benefits of testing so they can decide if they want to be tested” (American Cancer Society, 2017a) and the USPSTF’s online recommendation summary states, “The decision to start screening…should be an individual one. [Those] who place a higher value on the potential benefit than the potential harms may choose to begin [screening]” (U.S. Preventive Services Task Force, 2016). While we are aware that a general bias toward screening exists in the population, this does not ensure that every patient will necessarily be pro-screening. Thus, it is of the utmost importance that we understand what drives patient preferences for screening in the first place. Then, once these preferences are defined and understood, the information regarding the various positive and negative attributes of options—as well as the likelihoods associated with these attributes for an individual—and any other features of the decision could be tailored to the individual.

**Understanding Preferences for Screening**

In order to support patient decisions about screening, it is important for researchers and healthcare workers to understand what preferences an individual may have when evaluating a screening test. Previous work has elicited these preferences using both qualitative and quantitative methods. We posit here that the additional use of an individual difference measure of preferences for features of screening tests should be developed.

**Qualitative Measures of Preference**

To understand what information influences the cancer screening decision process, many have begun by simply asking patients what types of things matter most to them when deciding to be screened or which screening test to take. When considering whether to screen or not to screen, when to begin screening, when to stop screening, and how frequently to screen, many factors may be at play and it is especially important to understand these factors in an area where screening is frequently not recommened, such as PSA screening. Rai and colleagues (2007) began to identify some of these factors when they interviewed men who had raised the issue of being tested for prostate cancer with their physicians. Prior to meeting with their physicians, individuals explained why they wanted the test. Generally, these men stated they were interested in receiving the PSA test because they wanted information or believed diagnosis would increase their chances of being able to treat and survive the cancer, had friends or familiy with prostate cancer, or had heard about the prevalence of prostate cancer in the news. Interestingly, the authors report that,

Most men had not placed much importance on information about the benefits and limitations of testing in their decision to have a PSA test. They had been happy to receive some information from their GP, but they had already made their minds up to do something to alleviate their concerns about prostate cancer by the time they saw their GP, and this desire to ‘do’ something had remained unaltered. (Rai et al., 2007, p. 367)

This work demonstrates that the decision to be screened in the face of risk and recommendations to the contrary is determined by a multitude of factors beyond the risks and benefits of the test and even the recommendations of their physicians.

Other screening tests, such as those for colon cancer, have the opposite problem—fewer individuals are being screened than would be recommended. In this case it is also important to understand what factors may differ between those who would choose to be screened and those who would not. Gyrd-Hansen and Søgaard (2001) surveyed individuals regarding colorectal cancer screening. When investigating differences between those who wanted to receive tests and those who did not, those who wanted the test cited the importance of gaining information, being worried about cancer, wanting to reduce their chance of dying from cancer, potential regret if they later find out they have cancer, and a family history of cancer. Those who did not want to be screened listed concern over the result of the test, prefering to remain ignorant, believing the risks outweighted the benefits or that the test was ineffective, and worry about the cost of the test.

In realted work, one study (Greiner, Born, Nollen, & Ahluwalia, 2005) focused on colorectal cancer screening for low-income, African Americans patients. Qualitative analyses of participant responses indicated that six overarching themes were present. One factor bolstering the desire for screening was hope—the hope that screening could lead to early diagnosis and live saved. A second factor helping patients decide was accuracy—most of the participants opted for the most thorough, accurate screening test. Next there were four barriers to screening that were routinely vocalized: 1) fear that something may be wrong but not wanting to know, 2) mistrust of physicians and the system, 3) a lack of knowledge coupled with a desire to learn about the screening tests, and 4) fatalistic thinking that if cancer were to be found nothing could be done about it.

It is worthwhile here to note that for some cancers there are multiple screening tests to be considered. For instance, there are currently 7 different screening tests for colorectal cancer, each with varying accuracies, risks, and benefits (American Cancer Society, 2017b). This variety allows researchers to study not only the differences in those who choose to screen versus not screen, but also what influences their decision to choose one screening test over another. Themes mentioned previously are echoed in this type of work by Palmer and colleagues (2010) who focused on differences in which of two screening tests for colorectal cancer participants were interested in receiving. After completing in-depth personal interviews of participants, they reported that participants with a preference for colonoscopy cited the importance of thoroughness, accuracy, and efficiency as their rationale, while those with a preference for fecal occult blood test (FOBT) stated that factors such as simplicity, low risk or discomfort, and fear of sedation/hospital/clinic drove their decision. These same sentiments have been seen in other work using qualitative responses as well (Gyrd-Hansen & Søgaard, 2001; Wolf, Basch, Brouse, Shmukler, & Shea, 2006). Work by Wolf and colleagues (2006) interviewed individuals over the phone regarding colorectal cancer screening preferences (FOBT or colonoscopy) and their reasons for choosing that test. Individuals with no test preferences cited they would want their physician’s recommendation before choosing; those who preferred the FOBT listed the convenience, noninvasiveness, lack of pain, familiarity, their family’s preference, cost, and lack of embarrassment; and those who preferred colonoscopy referenced the test as already being recommended by their physician and being the most accurate.

**Quantitative Measures of Preference**

While these qualitative responses begin to give researchers an idea of what individuals value, it does not necessarily give insight into what trade-offs an individual would be willing to make, or what factors prove the most important at the population level. These types of questions can begin to be answered by investigating preferences through stated preference methods such as conjoint analysis and discrete choice experiments (experimental and computationally complex methods used to estimate the utility—or satisfaction—associated with a particular choice). These methods aim to describe each feature of the decision with a number of levels. Within the cancer screening literature one of the most common features is the accuracy of the test. Some studies depict accuracy as the sensitivity of the test (“Screening test is able to detect if you actually have” colorectal cancer in x/4 cancers; Nayaradou, Berchi, Dejardin, & Launoy, 2010, p. 226), while others describe it much more vaguely as accuracy (“Accuracy for finding cancer” is x%; Hawley et al., 2008). By systematically varying the values of x (for this and other features) and having participants respond to this set of options population-level preferences can be estimated.

For example, a study conducted by Nayaradou and colleagues (2010) participants showed a preference for the accuracy of the test (sensitivity) consistent with the qualitative work by Greiner and colleagues (2005). But participants showed no particular preference when it came to features such as the potential that screening would lead to additional testing that would ultimately be unnecessary. The work noted above by Gyrd-Hansen and Søgaard (2001) also utilized a discrete choice experiment. Their work focused on cost, test frequency, risk of false positive, and risk reduction, and found that for breast cancer risk reduction, risk of false positives, and frequency of screening were important to the decision process, while for colorectal cancer risk reduction and cost were the most important deciding factors. Beyond attempting to measure what general or population-level preferences may exist, other researchers have used these methodologies to measure preferences with the intent to predict screening intentions (Hawley et al., 2008; Imaeda, Bender, & Fraenkel, 2010). Work by Hawley and colleagues (2008) examined preferences for test accuracy, preparation process for the test, test discomfort, test frequency, test invasiveness, and how much of the cost insurance would cover. The authors reported that participants with a higher stated preference for accuracy were more likely to rank colonoscopy highly, whereas those who ranked discomfort or test frequency as important were more likely to rank FOBT highly. These results again echo the sentiments from the qualitative literature regarding how preferences for test attributes such as accuracy and comfort can influence decision-making (Phillips, Van Bebber, Marshall, Walsh, & Thabane, 2006).

**Individual Difference Measures**

While these qualitative and quantitative methodologies give insight into the overarching population-level preferences for features of a screening test, this does not mean that these preferences are equally strong in all individuals. Therefore, this generality does little to inform our predictions about how individuals may behave or how best to tailor information to improve the decision making process for these individuals. Further, it is possible that individuals that have the same experiences have different preferences. For instance, in the alcohol abuse literature, having a parent who abuses alcohol is considered a risk factor for most children, but for a subset, this is a protective factor (Rice, Dandreaux, Handley, & Chassin, 2006). Individual difference measures can begin to explain why individuals with the same background may have radically different preferences, and this may in turn assist us in tailoring risk communication to these people. Additionally, while many other individual difference measures exist which may measure similar attributes (e.g. the Cancer Anxiety Scale; Trumbo, McComas, & Kannaovakun, 2007; the Multidimensional Inventory of Hypochondriacal Traits; Longley, Watson, & Noyes, 2005;, the Value of Health scale; Lau, Hartman, & Ware, 1986 ) none of these are specifically targeted to identify differences in the importance of various attributes of screening tests. This information would ultimately assist in tailoring information to an individual’s needs. In order to tailor information to the individual, researchers need to first understand what individual differences exist that may be useful, then how to measure these individual differences. Once a measure is developed it can begin to answer questions and help inform health communications.

For example, measuring an individual’s desire to utilize medicine can help inform our predictions about how individuals will behave when making medical decisions. This is why Scherer and colleagues (2016) created the Medical Maximizer-Minimizer Scale. They were able to define and measure a unique feature of an individual, and to answer the question: “To what extent is this person a medical maximizer/minimizer?” This scale has shown that individuals who are more maximizing reportedly take more prescription medications on a daily basis, visit health care professionals more frequently, stay in the hospital more frequently, and even get the flu vaccine more frequently than minimizers (Scherer et al., 2016). This scale has been able to predict that maximizers are more likely to want to receive an MRI for migraine headaches when no other neurological symptoms are present (Scherer et al., 2016), and treatment when incidental findings are identified with diagnostic imaging tests (Kang et al., 2018).

**Need for a New Individual Difference Measure**

Given the overenthusiasm for some screening tests (such as PSA), the under-enthusiasm for other tests (such as colon cancer screening), and the increased number of screening tests, many individual differences seem to exist in screening decisions. From the qualitative and quantitative work above a number of features seem to hold consistent importance in screening test decisions. However, these works largely discuss these in terms of which preferences were most important, and do not discuss how individuals may vary in their preferences. Therefore, I propose that a new individual difference measure should be developed which focuses on the attributes of screening tests that individuals find important when making their decisions (both for whether to screen at all or which test to choose). The proposed measure would help in five ways.

First, this process will identify factors that are important to many patients when making their screening decisions. As discussed above, many of these factors have already been identified and will be utilized. These will be used as a starting point to determine what is important to patients when considering screening.

Second, this scale will allow us to identify the variability of these preferences. For instance, is the accuracy of a test important to all individuals equally, or do some people not find accuracy to be an important factor? By assessing people’s preference on many features we will be able to not only discuss what they find important, but how much the population varies in how important they believe these features to be.

Third, this scale will allow us to predict screening behaviors in the future based on the features of the screening test at hand and the preferences for these features an individual expresses. Within the domain of research much work has attempted to predict the decision to screen. Thus far, use of demographics, health histories, and existing scales have only been able to explain a proportion of the variance we see in screening decisions. A large amount of variability still exists that warrants explanation. This scale could additionally aid researchers in the ability to predict screening test uptake, as well as screening test choice in the future. This would lead to a better understanding of the individual, and better empirically tested interventions to inform individuals and tailor this screening information in a meaningful way.

Fourth, this scale will assist us in tailoring information to the individual patient in order to increase adherence to screening guidelines. These tailored messages are vital especially considering work which has shown that physician recommendations to patients tend to more closely match their own preferences than the preferences of their patients for colorectal cancer screening decisions (Ling, Moskowitz, Wachs, Pearson, & Schroy, 2001). Discordance has also been shown between patient preferences and colorectal cancer screening test completion (Hawley et al., 2012; Schroy et al., 2011) which suggests that patient’s preferences are not being appropriately incorporated into the decision process, ultimately leading to lower uptake of beneficial screening tests (Hawley et al., 2012). If patient preferences can be better understood and information can be better personalized, this could improve communication between patients and physicians and encourage better screening test adherence (Hawley et al., 2008).

Finally, this scale will be compared with other methods of preference elicitation. Much work has looked at measuring preferences (see Mansfield et al., 2016 for a review). However, no work to our knowledge has set about to compare different method of preference elicitation. While an individual difference scale (which we aim to develop here) is one way to directly elicit preferences, other methods exist as well. Indirect methods such a Discrete Choice Experiment (DCE) or a Threshold Technique (TT) ask more indirect questions of participants in an attempt to identify more global preferences for features (Stafinski et al., 2015). It is necessary to compare and contrast how these methods measure preferences, as well as how they are capable of predicting choices, in order to identify which preference elicitation methods may be worthwhile.

To begin to develop a scale capable of these aims, an instrument will be designed herein focusing on individuals’ preferences for attributes of screening tests by creating a set of possible factors regarding screening test attributes (Chapter 2), running an exploratory factor analysis to identify the underlying structure of these items (Chapter 3), completing a confirmatory factor analysis to confirm the structure of the items (Chapter 4), showing the scale is reliable (Chapter 5), has convergent and discriminant validity (Chapter 6), and investigate if the structure can be replicated in a more diverse population (Chapter 7). Additionally, we will compare this scale with other methods of measuring preferences (Chapter 8). Finally, we will conclude with a summary of the results found and their implications (Chapter 9).

**Chapter 2**

**Item Development**

Initially, the aim of this project was to create two scales. The first scale would look into preferences for attributes of screening tests for diseases in general (e.g. prenatal screening for Down Syndrome), while the second scale would be focused on preferences for attributes of screening tests specifically for cancer. As the first step in developing these questionnaires, a convenience literature review was completed by searching Google Scholar for 100 articles with the general search words, “screening test” preferences. Articles were coded for a number of variables including, but not limited to, the independent and dependent variables measured, and potential diagnosis (e.g. colon cancer, Down syndrome). These articles included both quantitative and qualitative assessments of patients, physicians, proxy decision-makers, and the general population’s reactions to screening tests, both for cancer, as well as for disease more broadly. A full guide to how the search was directed and what variables were coded for can be found in Appendix I, and the full literature review spreadsheet on can be explored at [osf.io/agx3p](file:///C:\Users\shafferv\Library\Group%20Containers\3L68KQB4HG.group.com.readdle.smartemail\databases\messagesData\1\25730\osf.io\agx3p). These articles were investigated and re-coded to identify different attributes of a screening test or individual differences of the participants that were hypothesized to play a role in screening choices (this can be seen in the second sheet of the excel file at [osf.io/agx3p](file:///C:\Users\shafferv\Library\Group%20Containers\3L68KQB4HG.group.com.readdle.smartemail\databases\messagesData\1\25730\osf.io\agx3p)). A total of 47 unique attributes were identified. The top 10 attributes were as follows: discomfort, cost/insurance, accuracy, time/duration of test, decision making preferences, frequency, complications, inconvenience/sedation, location, and comparative risk/risk.

Items for our scale were designed to match the attributes recovered from the literature review to maximize face validity. Based on the initial literature review 91 scale items were created that fell under 9 broad factors including test accuracy (accuracy), test safety (discomfort and complications), health benefits (the reverse of comparative risk/risk), cost (monetary cost/insurance, inconvenience/sedations, location, time/duration of test), familiarity/simplicity (including frequency), disease attributes, recommendations, personal fit/preferences, and desire for action. An implicit valance assumption was made when creating these items; all items were written in a positive format such that greater values indicated greater importance of that particular attribute. Thus, no items were created which required reverse coding. Test accuracy consisted of 11 items (e.g. “Gives me accurate information about whether or not I have the disease”), test safety consisted of 15 items (e.g. “Has no risk of pain”), health benefits consisted of 13 items (e.g. “Will improve my health”), cost consisted of 12 items (e.g. “Is cheap”), recommendations consisted of 12 items (e.g. “Is recommended by my physician”), familiarity/simplicity consisted of 10 items (e.g. “Is familiar”), personal fit/preferences consisted of 5 items (e.g. “Fits with my values”), action consisted of 6 items (e.g. “Will make me feel prepared”), and disease attributes consisted of the final 7 items (e.g. “Is for a preventable disease”). Items were identical for both “cancer” and “disease” scales. The full list of items may be viewed in Appendix III.

Next, the response scale was chosen. Each block of items was designed to be proceeded by the phrase, “It is important to me that the screening test I choose…”, and then each item was structured to be the completion of that sentence. For instance, one item may say, “It is important to me that the screening test I choose…Will improve my health”. Participants would be asked to respond on a 0-6 scale with endpoints of 0-This is the LEAST important thing to me and 6-This is the MOST important thing to me. One initial concern was that individuals would not properly utilize the entire scale given to them, and instead would rate all attributes as highly important (a screening test must have accuracy, benefit, be inexpensive, etc., essentially creating all negatively skewed items). To address this concern an introductory paragraph was created to encourage participants to use the entire scale. This paragraph also allowed us to alter the labeling of the screening test as being for “cancer” or for “disease”. This permitted us to see if these labels had any impact on the structure of responses of the two versions of this introduction. These introductions read as follows:

*When answering the following questions, we would like you to think about choosing a screening test for [cancer/disease]. We would like you to think about your own preferences toward [cancer/disease] screening and answer the questions as honestly as you can. Although many of these attributes may seem very important to you, we encourage you to use the whole scale available when responding. There are sometimes many tests when it comes to screening for [cancer/disease], and each test option may have different attributes that need to be considered to choose the best one for you.*

This introduction was altered to describe either a screening test for “disease” or a screening test for “cancer”. As previous work has noted that individuals prefer more invasive, surgical options (Omer, Hwang, Esserman, Howe, & Ozanne, 2013) and have stronger, more negative reactions (Nickel et al., 2015) when a diagnosis is labeled as “cancer”, it was worthwhile to investigate if the labeling of a screening test as being for cancer, or being for diseases more broadly, would impact responses.

**Chapter 3**

**Exploratory Factor Analysis**

*Rationale*

This study aimed to explore the possible structure underlying these concepts. In order to address this objective multiple exploratory factor analyses (EFAs) were conducted for the 91 items described in Chapter 2.

*Participants and Methods*

383 undergraduate students enrolled in an introductory psychology course at a Midwestern University were recruited to participate in this study for course credit. Individuals were asked demographic questions, and then were asked to complete the 91 items described in Chapter 2. Demographic information for all studies can be found in Appendix II. Items needed to be manually broken down into smaller blocks of items for ease of completion and presentation because participants were asked to respond to many items (184 items total: 91 items under the “cancer” heading, 91 items under the “disease” heading, and two attention check items, described below), and the online survey software utilized did not have the ability to do this for us. To break down items, all items were first randomized (so they were no longer in the order presented in Appendix III), then placed into blocks of ~19 items each (i.e. the first 19 items that appeared in the random order were put into block 1, the next 19 items were put into block 2, etc.). Within one block an attention check question was placed which stated, “Please choose the response labeled 1 on this item”. These blocks were presented in a random order, and the order the items were presented in within each block was also randomized.

*Analysis*

All analyses were conducted in R version 3.4.3 or later (R Core Team, 2017). Analyses for both the “cancer” items and the “disease” items separately revealed similar underlying structures. As such, only the “cancer” items and analyses will be discussed here. Tables of item descriptive statistics, model fit, and the final reduced scale for the disease items can be found in Appendix IV.

The underlying structure of the items was explored using multiple EFAs utilizing Oblimin[[1]](#footnote-1) rotation, which allowed for factors to be correlated. Simple structures were achieved by an iterative item reduction following the method proposed by Buchanan, Valentine, and Schulenberg ( 2014). Items were removed from further iterations if they did *not* load on a factor with values more extreme than -.3 or .3, loaded on more than one factor, or—for second iterations onward—switched which factor it loaded on. Items were removed in large batches (i.e. all items at the first iteration that did not load on any factor or loaded on more than one factor were removed prior to the second iteration being analyzed) in order to quickly reduce the scale to a manageable number of items. The final scale needed to be both short enough to be easily administered, but also contain enough items on each factor to ensure a proper fit. Thus, item reduction continued until all factors were composed of unique items and all factors had at least 3 items.

In order to compare these models we evaluated model fit indices. These indices included the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI), the Root Mean Squared Error of Approximation (RMSEA), and the Root Mean Squared Residual (RMSR). While higher values closer to 1 indicate better fit for CFI and TLI, lower values closer to 0 indicate better fit for RMSEA and RMSR (Tabachnick & Fidell, 2007).

*Results*

Data were screened and any individuals with missing data on the 91 items of interest (N=8) were removed. While 20% (N=75) of our sample incorrectly responded to the attention check item, this had no extensive effect on the final EFA solution, and thus they were allowed in the final dataset in an attempt to increase the power of the analysis. This resulted in a final dataset of 375 individuals. Data summary statistics were then viewed to ensure adequate range, mean, standard deviation, and skew for analysis and found to be acceptable. A table of these descriptive statistics is included in Appendix VI.

When attempting to discern the appropriate number of factors, parallel analysis (Ledesma & Valecro-Mora, 2007) and the Minimum Average Partial (MAP) criterion (Velicer, 1976) suggested 8 factors, the Kaiser criterion (Costello & Osborne, 2005) suggested 7, while the scree plot suggested 3, and the Very Simple Structure (VSS) criterion (Revelle & Rocklin, 1979) suggested 2 (see scree and VSS plots in Appendix VII). As our survey was designed with 9 factors in mind, anything up to a 9-Factor solution would have been permissible. Therefore, 8 separate EFA models were analyzed—ranging from an 8-Factor model to a 1-Factor model—in order to compare them. However, the 8- and 6-Factor EFAs resulted in too few items loading on the last factor. The 7-Factor EFA failed to converge at the second iteration. The 3-Factor model splintered at the second iteration showing very poor consistency (i.e. the majority of items that had loaded on one factor in iteration 1 were seen to load on two separate factors in iteration 2, making the resulting loadings largely uninterpretable). These models were therefore removed from consideration.

*Results*

The 1-Factor model had poor fit and the singular factor was largely uninterpretable. The 2-Factor model, which had better fit, was quite general, and the incredibly broad factors (with one factor highlighting how quick, easy, convenient, and familiar the test was, and the second factor focusing on the accuracy and benefit of the test) did not give us the distinction between attributes that we were aiming for in this scale. While the 4-Factor model had promising fit and understandable factors, these factors seemed broader when compared to those of the 5-Factor model (namely, the factor of cost included many other types of cost, including time, convenience, etc.). Finally, the 5-Factor solution was shown to fit well, and have clearly interpretable factors. Given these model fits and interpretability, our data are likely structured according to a 5-factor model. All fit indices, number of iterations, and the final number of items after item reduction are included in Table 1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 1. Model statistics for EFAs | | | | | | |
| Factor Model | CFI | TLI | RMSEA (90% CI) | RMSR | Iterations | Total Items |
| 1-Factor | 0.484 | 0.470 | 0.094 (0.088, NA) | 0.13 | 2 | 85 |
| 2-Factor | 0.749 | 0.734 | 0.068 (0.062, NA) | 0.05 | 3 | 78 |
| 4-Factor | 0.832 | 0.811 | 0.059 (.053, NA) | 0.04 | 3 | 76 |
| 5-Factor | 0.865 | 0.84 | 0.057 (.051, NA) | 0.04 | 4 | 67 |
| Reduced 5-Factor | 0.961 | 0.938 | 0.055 (0.046, 0.061) | 0.02 | 1 | 26 |
| Note: NAs for RMSEA upper limits are thought to arise because residuals are not distributed according to a noncentral chi square distribution. | | | | | | |

The 5-factor model could be reduced to 26 items, which are listed, along with their standardized loadings, in Table 2. Items were removed if they did not load highly on the factor (loading highly was determined as standardized loadings greater than .65), and all factors were assessed to ensure that at least 3 items loaded on each factor. The 5 factors can be interpreted as relating to the benefit of the screening test, the accuracy of the screening test, how familiar they are with the screening test, the monetary cost of the test, and how the test fits with their personal preferences.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table 2. Reduced 5-Factor, 26-item structure loadings | | | | | |
| Factor | Item | Standard-ized Loading | h2 | u2 | com |
| Accuracy  α=.93 | 1. Correctly classifies people who truly do NOT have the disease | 0.810 | 0.630 | 0.366 | 1.000 |
| 2. Is accurate at identifying people who truly have the disease | 0.810 | 0.660 | 0.339 | 1.000 |
| 3. Gives me accurate information about whether or not I have the disease | 0.800 | 0.650 | 0.354 | 1.000 |
| 4. Will not falsely diagnose me as having the disease if I do not have the disease | 0.780 | 0.540 | 0.457 | 1.000 |
| 5. Will tell me that I have the disease if I do, in fact, have the disease | 0.770 | 0.620 | 0.381 | 1.000 |
| 6. Will accurately determine the severity of the disease if it is detected | 0.730 | 0.610 | 0.387 | 1.000 |
| 7. Will give me consistent results (i.e. if I have the disease and am tested for it twice, both tests will tell me that I have the disease) | 0.720 | 0.600 | 0.396 | 1.000 |
| 8. Distinguishes between diseased and non-diseased people | 0.710 | 0.510 | 0.491 | 1.100 |
| 9. Is trustworthy | 0.640 | 0.640 | 0.363 | 1.200 |
| Benefit  α=.91 | 1. Will increase my life expectancy | 0.870 | 0.680 | 0.325 | 1.000 |
| 2. Will help me to live a longer life | 0.860 | 0.740 | 0.261 | 1.000 |
| 3. Will save my life | 0.790 | 0.620 | 0.378 | 1.000 |
| 4. Will keep me from dying | 0.750 | 0.570 | 0.426 | 1.000 |
| 5. Will reduce my chances of dying from the disease | 0.700 | 0.620 | 0.375 | 1.100 |
| 6. Will improve my health | 0.630 | 0.630 | 0.372 | 1.200 |
| Familiarity  α=.84 | 1. Is for a disease that I hear about in the news | 0.770 | 0.630 | 0.374 | 1.100 |
| 2. Is for a disease I have heard of | 0.720 | 0.560 | 0.445 | 1.100 |
| 3. Is encouraged by someone on TV | 0.720 | 0.580 | 0.419 | 1.100 |
| 4. Is a test my family member of friend has had before | 0.650 | 0.520 | 0.479 | 1.100 |
| 5. Is suggested by my child | 0.610 | 0.420 | 0.582 | 1.200 |
| Monetary Cost  α=.86 | 1. Is cheap | 0.860 | 0.710 | 0.289 | 1.000 |
| 2. Is affordable | 0.790 | 0.710 | 0.293 | 1.000 |
| 3. Is not expensive | 0.780 | 0.660 | 0.343 | 1.100 |
| Personal Preference  α=.81 | 1. Fits with my values | 0.970 | 0.920 | 0.081 | 1.000 |
| 2. Fits with my religious or spiritual beliefs | 0.630 | 0.540 | 0.463 | 1.200 |
| 3. Is consistent with my personal preferences | 0.550 | 0.480 | 0.521 | 1.300 |

*Conclusions and Limitations*

Through the use of multiple EFAs, several factor structures were assessed to understand how individuals might weigh the importance of various attributes of a cancer screening test. Evidence was found suggesting that there are 5 general factors that individuals attend to when deciding on a screening test: benefit, accuracy, familiarity, cost, and personal preference. Based on this structure and the purpose of the scale, this instrument was named “Evaluating Screening tests for Cancer: Attributes Patients Emphasize” (ESCAPE)**.** In order to confirm this structure, an additional sample will need to be collected and that data will need to be processed with a confirmatory factor analysis.

Limitations of this study include the sample of undergraduate students, who may not have sufficient exposure to screening tests to respond in a consistent manner to more diverse populations.

**Chapter 4**

**Confirmatory Factor Analysis**

*Rationale*

Seeing that both the “disease” and “cancer” items followed similar structures, the “disease” introduction and items were dropped in order to focus on the “cancer” items. All further studies focus solely on the “cancer items” and utilize the “cancer” labeled instructions. To confirm the factor structure assessed in the EFA study, data was collected from a new set of participants. Using this data, a confirmatory factor analysis (CFA) was completed.

*Participants and Methods*

All introductory psychology students were recruited to completed a mass pre-test survey online as part of a course requirement. Demographic information for all studies can be found in Appendix II. The mass pre-test contained over 400 items, 26 of which were our scale items. 852 students completed the 26 items of our scale and were included in the confirmatory factor analysis.

*Analyses*

The CFA was conducted using R version 3.4.3 (R Core Team, 2017) and the lavaan package (version 0.5-23.1097; Rosseel, 2012) for the 26 ESCAPE items following methods laid out by Buchanan, Valentine, and Schulenberg (2014). Again, fit indices were used to compare the various models. Along with CFI, TLI, RMSEA, and SRMR, additional indices including the Model Chi-Square (chisq), the Normed Fit Index (NFI), the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC). A model is preferred if it has a nonsignificant Chi-Square (though this is routinely violated when sample size is large; Byrne, 2016) and large values for NFI (close to one; Byrne, 2016). Both the AIC and BIC values are measure of comparative fit; the model with the lowest AIC/BIC value is preferred (Byrne, 2016).

*Results*

The model with 5 factors had good fit. Fit statistics are present in Table 3. Correlations among the 26 items for the reduced 5-factor model are shown in Figure 1. The upper triangle of the matrix shows the strength of the correlation by making larger correlations darker in color and showing direction of the correlation by changing the color from red (negative correlation) to blue (positive correlation). The lower triangle of the matrix indicates the actual values for each correlation, also altered by size and direction with color and opacity, respectively. Latent variable correlations are shown in Table 4. The latent variables (as measured by summary scores) predicted responses for each of their items (all p < .01) and are reported in Table 5.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3. Fit Indices for CFA | | | | | | | | |
| Parameters | chisq (289) | CFI | TLI | NFI | AIC | BIC | RMSEA (90%CI) | SRMR | |
| 62 | 1497.331, p<.001 | 0.93 | 0.921 | 0.915 | 63934.914 | 64229.264 | 0.07 (0.67-0.74) | 0.07 |

|  |
| --- |
|  |
| Figure 1. Correlations between scale items |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 4. Latent variable correlations | | | | |
|  | benefit | Accuracy | familiar | cost |
| accuracy | 0.711 |  |  |  |
| familiar | 0.106 | -0.121 |  |  |
| cost | 0.164 | 0.19 | 0.279 |  |
| fit | 0.306 | 0.21 | 0.565 | 0.228 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 5. Estimates and Standard Deviations of CFA Items | | | | | | |
| Item | Benefit Estimate (SE) | | Accuracy Estimate (SE) | Familiarity Estimate (SE) | Cost Estimate (SE) | Preference  Estimate (SE) | |
| 1. Will reduce my chances of dying from the disease | 1.00 (0.98) | |  |  |  |  | |
| 2. Will help me to live a longer life | 1.47 (0.06) | |  |  |  |  | |
| 3. Will keep me from dying | 1.04 (0.04) | |  |  |  |  | |
| 4. Will increase my life expectancy | 1.54 (0.06) | |  |  |  |  | |
| 5. Will save my life | 1.15 (0.04) | |  |  |  |  | |
| 6. Will improve my health | 1.22 (0.05) | |  |  |  |  | |
| 1. Will not falsely diagnose me as having the disease if I do not have the disease | |  | 1.00 (1.00) |  |  |  | |
| 2. Correctly classifies people who truly do NOT have the disease | |  | 0.95 (0.06) |  |  |  | |
| 3. Will give me consistent results (i.e. if I have the disease and am tested for it twice, both tests will tell me that I have the disease) | |  | 1.15 (0.05) |  |  |  | |
| 4. Gives me accurate information about whether or not I have the disease | |  | 1.08 (0.04) |  |  |  | |
| 5. Will accurately determine the severity of the disease if it is detected | |  | 1.23 (0.05) |  |  |  | |
| 6. Is accurate at identifying people who truly have the disease | |  | 1.53 (0.06) |  |  |  | |
| 7. Will tell me that I have the disease if I do, in fact, have the disease | |  | 1.62 (0.06) |  |  |  | |
| 8. Is trustworthy | |  | 1.00 (0.04) |  |  |  | |
| 9. Distinguishes between diseased and non-diseased people | |  | 1.06 (0.05) |  |  |  | |
| 1. Is encouraged by someone on TV | |  |  | 1.00 (0.88) |  |  | |
| 2. Is for a disease that I hear about in the news | |  |  | 1.21 (0.07) |  |  | |
| 3. Is suggested by my child | |  |  | 1.35 (0.09) |  |  | |
| 4. Is a test my family member of friend has had before | |  |  | 1.48 (0.09) |  |  | |
| 5. Is for a disease I have heard of | |  |  | 0.96(0.06) |  |  | |
| 1. Is cheap | |  |  |  | 1.00 (1.47) |  | |
| 2. Is not expensive | |  |  |  | 1.18 (0.04) |  | |
| 3. Is affordable | |  |  |  | 0.70 (0.02) |  | |
| 1. Fits with my values | |  |  |  |  | 1.00(1.05) | |
| 2. Is consistent with my personal preferences | |  |  |  |  | 1.04 (0.05) | |
| 3. Fits with my religious or spiritual beliefs | |  |  |  |  | 1.24 (0.07) | |

*Conclusions and Limitations*

This analysis suggests a 5-factor model is supported by the data. Limitations of this study include the sample of undergraduate students, who may respond differently to these items as compared to a more diverse population.

**Chapter 5**

**Assessing the Reliability of the ESCAPE Questionnaire**

*Rationale*

It is necessary to assess the reliability of any scale and to discuss the consistency with which the scale measures attitudes.

*Participants and Methods*

Undergraduate students (N=137) enrolled in an introductory psychology course at a Midwestern University completed our 26-item scale in this study for course credit. Demographic information for all studies can be found in Appendix II. These students were contacted two weeks later and asked to complete the 26-item scale for a second time. A total of 103 students completed both scales and could be identified at both time points.

*Analyses*

Raw Intraclass Correlation coefficients values (ICC2; Shrout & Fleiss, 1979) were calculated in R (R Core Team, 2017) using the *psych*  package (Revelle, 2018) and can be seen in Table 6. ICC2 value was used, instead of ICC3, as this initial gathering of evidence was thought to measure individuals who were thought of as randomly selected. If ICC3 were to be used, this would suggest that these individuals and preferences were instead fixed. ICC2 was chosen as it is the more generalizable of the ICC measures (Shrout & Fleiss, 1979). Factor structures were modeled and correlations between each factor at Time 1 and Time 2 were used to assess Tucker's congruence coefficient (TCC; Tucker, 1951) within Mplus (Muthén & Muthén, 2010). Factor scores showed the presence of many influential observations. The Mplus framework was utilized to identify outliers within our sample. We attempted to investigate why these individuals were flagged as outliers. By viewing these rows of data that were identified as influential, we found that these individuals were generally misusing the scale (e.g. using all 4s or using 1s and 4s only). As our instructions specifically stated that participants use the whole scale, these misuses of the scale seemed to be causing these points to be flagged as influential. Thus, we formulated a rule: rows were removed if they only utilized one or two values on our scale. Given this rule, 8 rows were removed and factor correlations were re-run. It is worthwhile to note that these values are from the individual subscale-subscale models (i.e. each subscale was modeled individually, as opposed to modeling the entire scale and then investigating subscales) due to our small sample size. When viewing these new models (with 8 outliers removed), an additional 4 rows were identified as influential that almost followed our rule (remove if using only one or two values of our scale), except that these rows also used a third scale value, but only once (e.g. using all 4s and 1s, but also one 2).

*Results*

Initial factor correlations were similar to ICC2 values. We can see that Tucker's congruence coefficients greatly improved with the removal of the first eight outliers. Further, we can see that when an additional four outliers were removed, comparison of the responses at Time 1 and Time 2 revealed acceptable relationships as can be seen in Table 6.

|  |  |  |  |
| --- | --- | --- | --- |
| Table 6. ICC 2 Estimates for Study 3 | | | |
| Subscale | ICC2 Raw | TCC w/o 8 | TCC w/o 12 |
| Accuracy | .56 | .63 | .64 |
| Benefit | .67 | .81 | .84 |
| Familiarity | .62 | .80 | .91 |
| Cost | .54 | .68 | .68 |
| Personal Preference | .61 | .86 | .87 |

*Conclusions and Limitations*

Analysis of two time points of data for the ESCAPE questionnaire suggest that the scale is reliable for the Benefit, Familiarity, and Personal Preference subscales when influential observations (individuals who did not utilize our full scale) were removed. However, reliability for the Accuracy and Cost subscales was less robust and should thus be interpreted cautiously when considering preferences over time. It is also worthwhile to note that given the small sample size at hand, these ICC values are subject to shrinkage that may not appear in a larger sample.

**Chapter 6**

**Assessing the Validity of the ESCAPE Questionnaire**

*Rationale*

It is important to assess how this new scale relates to previously validated scales that represent similar and dissimilar constructs.

*Participants and Methods*

389 undergraduate students enrolled in an introductory psychology course at a Midwestern University were recruited to participate in this study for course credit. Demographic information for all studies can be found in Appendix II. Participants responded to items about their demographics and personal history with cancer, and then were asked to complete a number of scales. Scales included our 26-item ESCAPE scale, healthcare access items (Pleis, Lucas, & Ward, 2009), the Beliefs about Medicine Questionnaire (BMQ-general; Horne, Weinman, & Hankins, 1999), the Domain Specific Risk Taking Scale (DOSPERT; Blais & Weber, 2006), a Cancer Anxiety Scale (CAS; Trumbo, McComas, & Kannaovakun, 2007), the Multidimensional Health Locus of Control scale (MHLC; Wallston, Wallston, Kaplan, & Maides, 1976), an Ambiguity Aversion in Medicine scale (AA-med; Han, Reeve, Moser, & Klein, 2009), the Subjective Numeracy Scale (SNS; Fagerlin et al., 2007), the Medical Maximizer-minimizer Scale (MMS; Scherer et al., 2016), the Multidimensional Inventory of Hypochondriacal Traits (MIHT; Longley, Watson, & Noyes, 2005), the Functional Assessment of Chronic Illness Therapy—Spiritual well-being scale (FACIT-Sp; Peterman, Fitchett, Brady, Hernandez, & Cella, 2002) with the 2 chronic illness items removed, the “Age-Universal” Religious Orientation I-E Scale (AURO-12; Gorsuch & Venable, 1983), a 3-item religiosity scale (Rostosky, Regnerus, & Wright, 2017), the Value of Health scale (VOH; Lau, Hartman, & Ware, 1986), the Health Opinion Survey (HOS; Krantz, Baum, & von Wideman, 1980), the 10 item Big 5 Inventory (Rammstedt & John, 2007), and the Internal External Control Scale (IECS; Rotter, 1966). The relationships hypothesized *a priori* are detailed in Appendix VIII.

*Analyses*

All data were analyzed using R (R Core Team, 2017). Correlations were analyzed for all variables where there was sufficient variance—all other variables have been left out of this discussion. All correlations, descriptions of measurement scales, and scale statistics are included in Table 7. It is important to note that the benefit and accuracy subscales are significantly negatively skewed (skew= -1.56 and -2.00 respectively), and thus their relationships with other variables may be altered by this non-normality. Appendix VIII compares the hypothesized vs. observed relationships between scales.

*Results*

Correlational analysis of the data revealed that the importance of accuracy was positively related to being white or female, being scared of cancer, being able to get the medicine you need, working well with numbers, being more aware of your health, and having more meaning in your life. It was negatively related to being more risk taking in healthcare, financial, and ethical situations and delaying a regular checkup because you cannot afford it.

The importance of benefit was positively related to being white, being worried or scared about cancer, being able to get the medicine you need, being a medical maximizer, being more aware of your health, having more meaning, peace, and spiritual well-being. It was negatively related to being risk taking in ethical situations and delaying a regular checkup because you cannot afford it.

The importance of cost was positively related to being worried about paying for an unexpected medical expense, avoiding going to see the doctor because you cannot afford it, getting more insurance if you could afford it, and desire to be screened in the future. The importance of cost was negatively related to being white. The importance of familiarity was not related to any of the included scales or items. The importance of personal preference was positively related to being female, being more averse to ambiguity, having more faith and meaning in life, having a more internal or external religious orientation, and being more religious.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 7. Measures, correlations, and descriptive statistics | | | | | | | | | | | | | |
| Measured Variable | Accuracy | | Benefit | | Cost | | Familiarity | | Personal Preference | | Measurement Scale  α reliability# | M (SD) | | |
| Age | 0.02 | | 0.03 | | -0.04 | | 0.01 | | -0.01 | | Number entry | 18.81 (1.447) | | |
| Gender  Females are more likely to rate accuracy and personal fit as important to them | 0.17\*\*\* | | 0.05 | | 0 | | -0.07 | | 0.14\* | | 0=male, 1=female | 61.70% female | | |
| Race  White participants are more likely to rate accuracy and benefit as important to them, whereas non-white participants are more likely to rate cost as important to them. | 0.12\* | | 0.12\* | | -0.15\*\*\* | | -0.1 | | 0 | | 0=nonwhite 1=white | 81.23% white | | |
| Subjective numeracy  Individuals with higher numeracy are more concerned with the accuracy of a test. | 0.22\* | | 0.18 | | 0.09 | | -0.07 | | -0.01 | | 1-6 scale  α=.80 | 4.220 (0.930) | | |
| Maximizer-Minimizer  Medical maximizers are more concerned with the benefit received from the test. | 0.14 | | 0.25\*\*\* | | -0.03 | | 0.09 | | 0.12 | | 1-7 scale  α=.77 | 4.394 (0.881) | | |
| BMQ harm | 0.16 | | 0.1 | | 0 | | -0.17 | | -0.12 | | 1-5 scale  α=.68 | 3.612 (0.375) | | |
| BMQ overuse | 0.01 | | 0.05 | | -0.06 | | -0.06 | | -0.05 | | 1-5 scale  α=.76 | 3.271 (0.451) | | |
| DOSPERT social | 0.05 | | -0.06 | | -0.08 | | -0.17 | | -0.14 | | 1-7 scale  α=.68 | 4.682 (0.791) | | |
| DOSPERT health  Individuals who are more risk taking in health are less likely to be concerned with the accuracy of the test. | -0.24\*\*\* | | -0.13 | | -0.08 | | 0.02 | | -0.12 | | 1-7 scale  α=.68 | 3.514 (1.216) | | |
| DOSPERT recreation | -0.08 | | -0.05 | | -0.01 | | -0.07 | | -0.01 | | 1-7 scale  α=.81 | 3.985 (1.469) | | |
| DOSPERT financial  Individuals who are more risk taking in finance are less likely to be concerned with the accuracy of the test. | -0.31\*\*\* | | -0.19 | | -0.17 | | 0.06 | | -0.05 | | 1-7 scale  α=.82 | 2.974 (1.320) | | |
| DOSPERT ethical  Individuals who are more risk taking in ethical situations are less likely to be concerned with the accuracy of the test or the benefit of the test. | -0.39\*\*\* | | -0.26\*\*\* | | -0.03 | | 0.13 | | -0.11 | | 1-7 scale  α=.76 | 2.598 (1.123) | | |
| Cancer anxiety scale | 0.04 | | 0.08 | | 0.03 | | 0.08 | | 0.1 | | 1-5 scale  α=.46 | 3.258 (1.033) | | |
| MHLC internal | 0.2 | 0.21 | | 0.05 | | -0.05 | | 0.02 | | 1-5 scale  α=.73 | | | 3.217 (0.891) | | |
| MHLC powerful others | -0.04 | 0.02 | | 0 | | 0 | | 0.07 | | 1-5 scale  α=.60 | | | 2.675 (0.979) | | |
| MHLC chance | -0.07 | | 0.01 | | 0.02 | | 0.03 | | -0.06 | | 1-5 scale  α=.59 | 2.745 (0.882) | | |
| AA-med  Those who are more averse to ambiguity are more likely to be concerned with personal fit. | 0.16 | | 0.15 | | 0.14 | | 0.05 | | 0.25\*\*\* | | 1-5 scale  α=.68 | 3.234 (0.656) | | |
| MIHT cognitive | -0.05 | | -0.01 | | 0.16 | | 0.12 | | -0.01 | | 1-5 scale  α=.92 | 3.484 (1.646) | | |
| MIHT behavioral  Those higher in the need for reassurance of their health status are more concerned with the benefit of the test. | 0.15 | | 0.22\* | | 0.05 | | 0.1 | | 0.15 | | 1-5 scale  α=.84 | 3.392 (1.021) | | |
| MIHT perceptual  Those who are more aware of their health are more concerned with the accuracy and benefit of the screening test. | 0.41\*\*\* | | 0.27\*\*\* | | 0.05 | | -0.07 | | 0.1 | | 1-5 scale  α=.83 | 3.904 (0.638) | | |
| MIHT affect | 0.07 | | 0.17 | | 0.02 | | 0.18 | | 0.13 | | 1-5 scale  α=.85 | 3.017 (0.873) | | |
| FACIT meaning  Higher scores in spiritual meaning were related to rating accuracy and benefit as more important. | 0.24\*\*\* | | 0.27\*\*\* | | -0.06 | | -0.03 | | 0.18 | | 0-16 score  α=.85 | 12.04 (3.756) | | |
| FACIT peace  Higher scores in feelings of peace were related to rating benefit as more important. | 0.06 | | 0.25\*\*\* | | 0.02 | | 0.1 | | 0.14 | | 0-16 score  α=.81 | 10.05 (3.100) | | |
| FACIT faith  Higher scores in faith were related to rating personal fit as more important. | -0.02 | | 0.03 | | 0.05 | | 0.12 | | 0.48\*\*\* | | 0-8 score  α=.97, *r*=.95 | 4.393 (3.432) | | |
| FACIT  Those with higher scores in spiritual well being were more concerned with the benefit of the test as well as how it fits with their personal preferences | 0.13 | | 0.24\*\*\* | | 0 | | 0.08 | | 0.32\*\*\* | | 0-40 score  α=.69 | 26.46 (8.874) | | |
| AURO internal  Those with more internal religious orientations were more concerned with personal fit. | -0.04 | | 0.01 | | 0.02 | | 0.07 | | 0.43\*\*\* | | 1-7 scale  α=.69 | 3.949 (1.682) | | |
| AURO external  Those with more external religious orientations were more concerned with personal fit. | -0.13 | | -0.03 | | 0.01 | | 0.12 | | 0.4\*\*\* | | 1-7 scale  α=.92 | 3.560 (1.596) | | |
| Religiosity  Those scoring higher on religiosity were more concerned with personal fit. | -0.02 | | -0.05 | | 0.05 | | 0.04 | | 0.42\*\*\* | | 1-3 scale  α=.88 | 2.185 (1.180) | | |
| Value of Health | -0.2 | | -0.06 | | 0.01 | | 0.04 | | -0.02 | | 1-4 scale  α=-.28 | 3.970 (0.726) | | |
| HOS information | -0.03 | | -0.05 | | 0.13 | | 0.08 | | 0.07 | | 0-7  α= | 3.567 (0.527) | | |
| HOS behavior | -0.03 | | -0.05 | | 0.13 | | 0.08 | | 0.07 | | 0-9  α= | 4.701 (1.581) | | |
| Big 5 openness | 0.14 | | 0.1 | | -0.02 | | -0.07 | | -0.11 | | 1-5 scale  α=.22, *r*= 0.13 | 3.433 (0.829) | | |
| Big 5 conscientiousness | 0.1 | | 0.07 | | -0.12 | | -0.14 | | 0.09 | | 1-5 scale  α=.29, *r*=0,17 | 3.484 (0.857) | | |
| Big 5 extraversion | 0.03 | | 0.08 | | -0.07 | | 0.02 | | 0.14 | | 1-5 scale  α=.58, *r*=.40 | 3.341 (0.651) | | |
| Big 5 agreeableness | 0.14 | | 0.15 | | 0.02 | | 0.03 | | 0.05 | | 1-5 scale  α=.33, *r*=.20 | 3.646 (0.866) | | |
| Big 5 neuroticism | 0.05 | | 0.01 | | 0.01 | | 0.03 | | -0.02 | | 1-5 scale  α=.51, *r*=.35 | 3.075 (1.024) | | |
| IE control | 0.03 | | 0.03 | | 0.05 | | 0.07 | | -0.02 | | Sum between 0-23  α=.59 | 12.1 (2.804) | | |
| Do you avoid or put off going to the doctor because you cannot afford it?  Those who more frequently put off visiting the doctor due to cost are also more concerned with the cost of a screening test | -0.04 | | -0.2 | | 0.29\*\*\* | | -0.02 | | 0.01 | | 1 (no, never)-7 (yes, all the time) | 2.026 (0.707) | | |
| How much does cancer scare you?  Those who are more scared of cancer are more concerned with the accuracy and benefit of the test. | 0.24\*\*\* | | 0.28\*\*\* | | 0.06 | | 0.07 | | 0.16 | | 1 (not at all)-7(extremely) | 5.437 (1.453) | | |
| When you think about the possibility of having cancer, how worried does it make you feel?  Those who are more worried by cancer are more concerned with the benefit of the test. | 0.2 | | 0.26\*\*\* | | 0.12 | | 0.13 | | 0.16 | | 1 (not at all)-7(extremely) | 5.026 (1.692) | | |
| How likely do you think it is that you will get cancer one day? | 0.18 | | 0.1 | | -0.03 | | 0.01 | | -0.03 | | 1 (not at all likely)-7(extremely likely) | 4.105 (1.732) | | |
| How difficult is it for you to find health insurance that you can afford? | 0.19 | | 0.18 | | -0.13 | | 0.08 | | 0.01 | | 1 (very difficult)-3 (not at all difficult) | 2.543 (0.333) | | |
| How difficult is it for you to find a health insurance plan with the type of coverage that you need? | 0.15 | | 0.13 | | -0.16 | | 0.05 | | -0.01 | | 1 (very difficult)-3 (not at all difficult) | 2.526 (0.333) | | |
| Would you get more medical care if you could afford it?  Those who state they would get more medical care if they could afford it are also more likely to state that the cost of a screening test is important to them. | -0.05 | | -0.01 | | 0.25\*\*\* | | 0.11 | | 0.05 | | 1 (no, not at all)-7 (yes, definitely) | 4.626 (2.121) | | |
| Can you get the medical care that you need?  Those who can get the medical care that they need are more likely to be concerned with the accuracy and benefit of a screening test. | 0.25\*\*\* | | 0.32\*\*\* | | -0.18 | | -0.07 | | 0.01 | | 1 (no, not at all)-7 (yes, definitely) | 6.307 (1.014) | | |
| If you get sick or have an accident, how worried are you that you will not be able to pay your medical bills?  Those who are more concerned they may not be able to pay their medical bills should they have an accident are also more concerned with the cost of a screening test. | -0.06 | | -0.1 | | 0.34\*\*\* | | 0.09 | | 0.12 | | 1 (not at all worried) -3 (very worried) | 1.532 (0.500) | | |
| How many first degree relatives (i.e. parents or siblings) do you have who have had cancer? | -0.03 | | 0 | | -0.03 | | 0 | | -0.01 | | 0-20+ | 1.43 (3.333) | | |
| How many second degree relatives (i.e. aunts, uncles, cousins) do you have who have had cancer? | 0.02 | | 0.04 | | -0.04 | | -0.02 | | -0.02 | | 0-20+ | 2.992 (2.667) | | |
| How many friends do you have who have had cancer? | -0.1 | | -0.02 | | 0.02 | | 0.1 | | 0.06 | | 0-20+ | 2.109 (0.928) | | |
| Health Access Item/Delay: Once you get there, you have to wait too long to see the doctor | -0.05 | | -0.07 | | 0.08 | | 0.04 | | 0 | | 0=no, 1=yes | 16.17% yes | | |
| Health Care Access Item/Afford: Getting a regular health check up  Those who have not had a regular health checkup because they could not afford it are less likely to be concerned with the accuracy and benefit of a screening test. | -0.18\*\*\* | | -0.15\*\*\* | | 0.04 | | 0 | | 0 | | 0=no, 1=yes | 12.04% yes | | |
| Have you ever been screened for cancer? | -0.06 | | -0.08 | | -0.01 | | 0.02 | | -0.07 | | 0=no, 1=yes | 5.14% yes | | |
| Would you want to be screened for cancer in the future?  Those who want to be screened for cancer in the future are more likely to rate accuracy and cost of the screening test as important. | 0.11\* | | 0.08 | | 0.12\* | | 0.03 | | 0.01 | | 0=no, 1=yes | 83.8% yes | | |
| Note: #reliability values for 2-item scales were calculated using both Cronbach’s Alpha (α), as well as Pearson correlations (*r*). | | | | | | | | | | | | | |

*Conclusions and Limitations*

The 5 subscales were generally correlated with other scales in the hypothesized direction. However, race and gender were unexpectedly related to the ESCAPE scale (as no predictions for these were made *a priori*), and the MMS scale was not related to any of the ESCAPE subscales.

Limitations include the sample utilized, especially as many of these items (the healthcare access items in particular) showed no variance and were therefore unable to be analyzed.

**Chapter 7**

**Scale Generalizability Study**

*Rationale*

This study aimed to determine if the factor structure assessed and validated within an undergraduate student population would generalize to a more diverse population. . Given that the initial sample was made up of undergraduate students, I believed it was critical that we first identify if this older, more diverse sample, had the same factor structure as the undergraduate sample. As this more diverse sample would likely have additional experiences with healthcare (e.g. paying for insurance, taking time off work to visit a physician, having a colonoscopy), I hypothesized that these older, more experienced individuals may have different priorities and preferences than our younger, less experienced sample previously utilized, which would result in different factor structures. As such, this more diverse sample was asked to complete all 91 of our original items.

In order to identify if this new sample held the same preference structure as the undergraduates, first, a multiple group EFA was conducted to identify where the two samples (undergraduates and Amazon Mechanical Turk workers) may be invariant; second, multiple exploratory factor analyses (EFAs) were conducted on new data collected from a more diverse sample (Amazon Mechanical Turk workers).

*Participants and Methods*

300 Amazon Mechanical Turk workers (Mturkers) were recruited to complete a survey online and were payed $1.15 for their participation. Demographic information for all studies can be found in Appendix II. The University of Missouri Department of Psychology’s Graduate Research Award provided funding for this project. Individuals responded to background and demographic questions. Participants were asked to completed the original 92 survey items (91 survey items + 1 attention check item). We felt it important to include all items, not just the pared-down scale shown previously, as we suspected that different factor structures may arise in a more diverse sample. Finally, participants were asked a number of questions regarding acceptability of false positives and false negatives.

*Analysis*

Multiple group EFAs were conducted in Mplus (Muthén & Muthén, 2010). Given that the EFA and CFA of the undergraduate sample identified a 5-Factor solution, the multiple group EFA tested whether a 5-Factor EFA using an oblique geomin rotation fit both the undergraduate and Mturk data similarly using methods laid out by (Muthén & Muthén, 2010). The first model run in this analysis imposed no equality constraints on the model parameters across the two groups. This model was then compared to the second model, where factor loading matrices are required to be invariant. A third model set both factor loading matrices and intercepts to be invariant. Further models restricted the variance/covariance matrices and factor means to be invariant. When comparing these models, as noted before, we look for large values for CFI and TLI, small values for RMSEA and SRMR, and the lowest value for BIC.

EFAs were conducted in R version 3.5.0 (R Core Team, 2017). To identify the underlying structure of the items for the Mturkers, multiple EFAs were conducted utilizing Oblimin[[2]](#footnote-2) rotation, which allows for factors to be correlated. Just as with the undergraduate sample in Chapter 3, simple structures were achieved by an iterative item reduction.

When attempting to discern the appropriate number of factors, parallel analysis, the MAP criterion, and the Kaiser criterion suggested 8 factors, while the scree plot suggested 3, and the VSS criterion suggested 2 (see scree and VSS plots in Appendix X). Again, as our survey was initially designed with 9 factors in mind, anything up to a 9-Factor solution would have been permissible. Therefore, 8 separate EFA models were analyzed—ranging from an 8-Factor model to a 1-Factor model—in order to compare them. However, the 8- and 7-Factor EFAs resulted in too few items loading on the last factor and were therefore removed from consideration.

In order to compare these models we evaluated model fit indices just as in Chapter 3. While higher values closer to 1 indicate better fit for CFI and TLI, lower values closer to 0 indicate better fit for RMSEA and SRMR.

*Results*

Data were screened for low-quality data prior to approving Mturker’s payments, resulting in a final dataset of 300 individuals. This low-quality data detection process is detailed in Appendix XI Data summary statistics were then viewed to ensure adequate range, mean, standard deviation, and skew for analysis and found to be acceptable as can be seen in Appendix XII.

First, to identify whether the undergraduate data and the Mturk data were invariant, a multiple group EFA was conducted. All model fits can be seen in Table 8. Given our data from the undergraduates and the Mturkers, the first (no invariance) model fit well, as did the second (factor loading invariance) model. However, the second model had slightly better BIC values. The third (factor loading and intercept invariance) model showed an appreciable drop in fit from the previous models, and further models including additional invariance of variance/covariances and factor means did not converge. Given these models and fits, the data supported a model that allowed the factor loadings to be set equal across the two groups, but beyond this, we see that these two groups were quite different from one another. As such, the data from the Mturkers should be analyzed separate from the undergraduate data to identify the underlying factor structure for this sample.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 8. Model fit for multiple group EFAs | | | | | | |
| Model | Number of Free Parameters | BIC | RMSEA (90% CI) | CFI | TLI | SRMR |
| 1: No equality constraints | 1254 | 246842 | 0.053 (0.052, 0.054) | 0.822 | 0.801 | 0.037 |
| 2: Equal Factor Loadings | 824 | 244861 | 0.053 (0.052, 0.054) | 0.812 | 0.801 | 0.050 |
| 3: Equal Factor Loadings and Intercepts | 734 | 245318 | 0.055 (0.054, 0.056) | 0.791 | 0.782 | 0.067 |

To identify the underlying structure of the items for the Mturkers, multiple EFAs were conducted. The 1-, 2-, and 3-Factor models had poor fit and their factors were largely uninterpretable. While the 4-, and 5-Factor models had promising fit and understandable factors, the 4-Factor models seemed to have broad categories, and the 5-Factor model seemed to contain five of the same factors as the 6-Factor model. Finally, the 6-Factor solution was shown to fit well and have clearly interpretable factors. These analyses suggest that the data from a more diverse sample are structured according to a 6-factor model. All fit indices, number of iterations, and the final number of items after item reduction is included in Table 9.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 9. Model statistics for EFAs | | | | | | |
| Factor Model | CFI | TLI | RMSEA (90% CI) | RMSR | Iterations | Total Items |
| 1-Factor | 0.50 | 0.481 | 0.091 (0.084, NA) | 0.12 | 4 | 70 |
| 2-Factor | 0.70 | 0.686 | 0.068 (0.06, NA) | 0.07 | 3 | 74 |
| 4-Factor | 0.83 | 0.807 | 0.060 (0.052, NA) | 0.05 | 3 | 63 |
| 5-Factor | 0.84 | 0.815 | 0.055 (0.046, NA) | 0.04 | 3 | 74 |
| 6-Factor | 0.86 | 0.838 | 0.052 (0.041, NA) | 0.04 | 3 | 80 |
| Reduced 6-Factor | 0.99 | 0.984 | 0.028 (0, 0.036) | 0.02 | 1 | 23 |
| Note: NAs for RMSEA upper limits are thought to arise because residuals are not distributed according to a noncentral chi square distribution. | | | | | | |

The 6-factor model could be reduced to 24 items, which are listed, along with their standardized loadings, in Table 10. Items were removed if they did not load highly on the factor (loading highly was determined as standardized loadings greater than .60), and all factors were assessed to ensure that at least 3 items loaded on each factor. The 6 factors can be interpreted as relating to how familiar the individual is with the screening test, the ease or comfort of receiving the screening, the accuracy of the test, the benefit of the test, a need to take action, and attributes of the disease that is being screened for.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table 10. Reduced 6-Factor 23 item loadings | | | | | |
| Factor | Item | Standardized Loading | h2 | u2 | com |
| Familiarity  α=.84 | Is encouraged by someone on TV | 0.79 | 0.69 | 0.31 | 1 |
| Is endorsed by an advertisement | 0.83 | 0.74 | 0.26 | 1 |
| Is a test my family member of friend has had before | 0.63 | 0.57 | 0.43 | 1.2 |
| Ease/Comfort  α=.90 | Will result in minimal discomfort | 0.87 | 0.71 | 0.29 | 1.1 |
| Will be a comfortable experience | 0.72 | 0.61 | 0.39 | 1.2 |
| Is convenient | 0.72 | 0.56 | 0.44 | 1 |
| Is a simple procedure | 0.71 | 0.61 | 0.39 | 1.1 |
| Has no risk of pain | 0.73 | 0.6 | 0.4 | 1 |
| Is quick | 0.7 | 0.58 | 0.42 | 1.2 |
| Accuracy  α=.85 | Is accurate at identifying people who truly have the disease | 0.66 | 0.57 | 0.43 | 1.1 |
| Will tell me that I have the disease if I do, in fact, have the disease | 0.75 | 0.59 | 0.41 | 1 |
| Will not falsely diagnose me as having the disease if I do not have the disease | 0.64 | 0.54 | 0.46 | 1.1 |
| Will give me consistent results (i.e. if I have the disease and am tested for it twice, both tests will tell me that I have the disease) | 0.6 | 0.46 | 0.54 | 1.1 |
| Will tell me that I do not have the disease if I, in fact, do not have the disease | 0.63 | 0.32 | 0.68 | 1.2 |
| Benefit  α=.84 | Will keep me from dying | 0.91 | 0.75 | 0.25 | 1 |
| Will help me to live a longer life | 0.64 | 0.6 | 0.4 | 1.4 |
| Will save my life | 0.72 | 0.66 | 0.34 | 1.1 |
| Action  α=.76 | Makes me feel that I am taking action toward better health | 0.92 | 0.84 | 0.16 | 1 |
| Will make me feel I have done something proactive | 0.57 | 0.41 | 0.59 | 1.3 |
| Makes me feel safer | 0.53 | 0.53 | 0.47 | 1.6 |
| Disease Attributes  α=.84 | Is for a curable disease | 0.94 | 0.83 | 0.17 | 1 |
| Is for a treatable disease | 0.74 | 0.6 | 0.4 | 1 |
| Is for a preventable disease | 0.65 | 0.57 | 0.43 | 1.1 |

*Conclusions and Limitations*

In this study, we conducted EFAs on data collected from a more diverse population than our initial sample of undergraduate students and found that the two samples respond slightly differently to these items. While some factors are similar between the two cohorts (e.g. accuracy and benefit), some factors are new (e.g. ease/comfort of test, need for action, and disease attributes). Also, while the familiarity factor in this sample consisted of three items, this subscale in the undergraduate sample contained two of these items (e.g. “encouraged by someone on TV” and “family member or friend has had before”). It could be argued that while these items are slightly different, they hold a similar ideology. For instance, in the undergraduate sample it seems that individuals want a test they have heard about (i.e. “heard about in the news”, “for a disease I have heard of”), similarly, the more diverse sample supported a similar item that lends itself both to something they have heard of, but also something that they have heard positive things about (i.e. “endorsed by an advertisement”).

Taken together, these analyses suggest that a more diverse population had a different underlying structure than our initial sample of undergraduates. Given this evidence, it is necessary to further confirm the structure of this scale within our more diverse population, and to give additional time and thought to why younger v. older adults may respond differently to these types of questions.

Limitations of this study include that the sample was not representative. While likely closer to the structure of the general population than our sample of undergraduates, this still represents a constraint of the current study. Additionally, this was solely an exploratory factor analysis. A further confirmatory factor analysis is still required.

**Chapter 8**

**Comparison of Stated Preference Methods Study**

*Rationale*

This study aimed to compare three methods of measuring preferences. The first of these methods utilizes the ESCAPE scale created in the previous chapters (for undergraduates). As the ESCAPE scale is meant to identify meaningful individual differences in the way people appraise—and thus make choices regarding—cancer screening tests, we would predict that responses on the 5 subscales of the ESCAPE scale would be predictive of their responses on items regarding cancer screening decisions.

The second method we used to measure preferences was a Discrete Choice Experiment (DCE). DCEs present an individual with two choices that vary on a number of attributes. Here, we varied choices based on 5 attributes: the benefit of the test in terms of lives saved, the risk of a false positive diagnosis, the risk of a false negative diagnosis, the familiarity of the test, and the cost of the test.[[3]](#footnote-3) An example item can be viewed in Figure 2. As one choice is not always preferable to another, participants are required to make trade-offs in order to choose between two choices. By carefully choosing the combination of attributes in each choice it is possible to calculate threshold values—the value up to which an individual is willing to accept a negative attribute in order to gain an increase of one unit in a positive attribute. Given that these thresholds inform us of the importance people place on the possible risks a screening test may hold, we would predict that these DCE threshold values would be predictive of their responses on items regarding cancer screening decisions.

|  |
| --- |
|  |
| Figure 2. Example of DCE item. |

The third method uses a threshold technique (TT) to identify the threshold at which an individual is willing to accept a negative attribute (or risk feature) in order to gain an increase of one unit in a positive attribute. For this experiment, our negative attributes consisted of the risk of a false positive, the risk of a false negative, and the cost of the screening test.[[4]](#footnote-4) Unlike the DCE, the TT asks participants an iterative set of questions to get to this threshold. For example, an individual may be asked to choose between two screening tests, one where the test saves 2 lives but has a 20% chance of a false positive diagnosis, and one where the test saves 1 life but has a 5% chance of a false positive diagnosis. If an individual is willing to accept high levels of false positives for the increase in lives saved, they should choose the first test. Then they would be asked to choose between two tests, one where the test saves 2 lives but has a 40% chance of a false positive diagnosis, and one where the test saves 1 life but has a 5% change of a false positive diagnosis. Thus, if an individual chooses a test with the higher benefit (lives saved) and a higher risk (false positive) the risk is increased until an individual is no longer willing to accept the benefit given the risk. As these thresholds also inform us of the importance people place on the possible risks a screening test may hold, we would predict that these TT threshold values will be predictive of their responses on items regarding cancer screening decisions.

Currently, methods such as the DCE and TT are utilized in a number of important ways in the literature. Both methods are routinely used to inform health policy and identify acceptable risks in treatments (Kopec et al., 2017; Ryan & Gerard, 2003). Additionally, these methods (and altered forms of them) are frequently included in values clarification exercises during shared decision making. These exercises help to engage the participant and clarify their preferences toward the decision at hand. Use of these exercises is thought to better prepare individuals to make the decision at hand, leading to less regret (Feldman-stewart et al., 2012).

To date, no study (of which we are aware) has compared the outputs of these different preference elicitation techniques. However, values from DCE and TT methods are currently utilized to make vital decisions about resource allocation in a variety of important settings including health care (Kennedy et al., 2011; Ryan, 2004) and health economics (Guttmann, Castle, & Fiebig, 2009). Thus, it is useful to identify any inconsistencies between these two methods both in terms of average thresholds within the sample and unique thresholds for the individual.

As these methods are thought to elicit the same type of information that an individual difference scale should measure, it is important to identify how similar these methods are and where they may differ. Additionally, many researchers have argued that preferences should not be seen as stable, and instead may be constructed at the time of elicitation, or altered by the mode of elicitation (Koopmans, 1964; Slovic, 1995; Tversky & Thaler, 1990). Therefore, a comparison of these multiple methodologies may help support or refute these ideas.

It is also worthwhile to identify which, if any of these methods is best at predicting screening intentions in individuals. As the purpose of the ESACPE scale is to be able to quickly administer the scale and identify preferences while in the clinical encounter, it would be important to be able to state that this method of measuring preferences is better in some way than DCE or TT that require more time and computational ability, both to build and to calculate values.

*Participants and Methods*

300 undergraduate students enrolled in an introductory psychology course at a Midwestern University were recruited to participate in this study for course credit. Demographic information for all studies can be found in Appendix II. All DCE items were created using Sawtooth software. Participants responded to items about their demographics and personal history with cancer, were asked to complete the DCE, the TT, and the items for the ESCAPE scale. Next, participants completed a number of scales including the Medical Maximizer-minimizer Scale (Scherer et al., 2016) and the Subjective Numeracy Scale (Fagerlin et al., 2007). Then, participants were told about a hypothetical test screening for breast/prostate cancer (matched to sex) which research has shown does not reduce the number of individuals who die from cancer (i.e. has no benefit) and were asked if they wanted this test. All materials can be found at [osf.io/xd9w6](file:///C:\Users\shafferv\Library\Group%20Containers\3L68KQB4HG.group.com.readdle.smartemail\databases\messagesData\1\25730\osf.io\xd9w6).

*Analysis*

All analyses were conducted in R version 3.5.0 (R Core Team, 2017). First, data from the DCE were used to create DCE threshold values calculated at the level of the sample, as well as the level of the individual. The TT data were used to calculate TT threshold values at the level of the individual. Information on how these calculations were conducted can be found in Appendix XIII. Next, ESCAPE subscales were created and summary scores were calculated. After this, a number of items were used to flag potential low-quality data. Information on how this data was flagged and removed can be found in Appendix XIV. When potentially low-quality individuals were removed, results varied slightly; see [osf.io/xd9w6](file:///C:\Users\shafferv\Library\Group%20Containers\3L68KQB4HG.group.com.readdle.smartemail\databases\messagesData\1\25730\osf.io\xd9w6) for details. Due to the small sample size and the exploratory nature of our study, we allowed all individuals into the final analyses.

As data for the DCE can be calculated two ways (one way which creates the sample level threshold and another that creates each individual’s threshold), data for both of these methods were compared to that of the TT method (which is only calculated at the level of the individual). It is important to compare the sample-level DCE estimate to the average individual-level TT estimate as these are the formats that are most commonly used for each method when discussing results and implications. Following this, individual-level DCE thresholds, individual-level TT thresholds, and individual’s ESCAPE subscales were correlated to investigate relationships.

Then, the DCE, TT and ESCAPE values were each predicted in their own linear regressions by age, being previously screened (0=no, 1=yes), desire to screen, worry about cancer, how scared they were of getting cancer, how likely they thought they were to get cancer, number of first degree family members with cancer, number of second degree family members with cancer, number of friends with cancer, sex (1=male, 0=female), race (1=white, 0=nonwhite), their Subjective Numeracy Score, and their Medical Maximizer Score. After this, we predicted participants’ intentions to screen via a hypothetical cancer screening test using the DCE thresholds, the TT thresholds, and the ESCAPE scale in a simultaneous logistic regression.

*Results*

DCE v. TT Threshold Values*.* When comparing thresholds from the DCE at the level of the sample to the TT methods, we find that estimates for false positives are similar, with TT values being slightly higher. Estimates for false negatives are similar, with DCE thresholds being slightly higher. Cost estimates vary considerably, with TT having a much higher threshold than DCE. See Table 11. However, when we instead look at thresholds calculated from the DCE at the level of the individual, these values greatly increase and are far less comparable to the TT estimates.

|  |  |  |  |
| --- | --- | --- | --- |
| Table 11. DCE and TT thresholds calculated at the sample-level and individual-level | | | |
|  | DCE for Sample-Level Data  M | DCE for Individual-Level Data  M | TT  M (SD) |
| False Positive | 17.33 | 2724.70 | 21.10 (24.63) |
| False Negative | 11.84 | 150.33 | 8.61 (8.45) |
| Cost | 120.79 | 5.30 e+12 | 842.59 (1091.26) |

Due to the method of calculating the DCE threshold at the level of the individual these estimates can become incredibly large. For instance, if an individual places a high utility on the benefit of a screening test saving two lives as compared to one life and places the same level of utility on a cost of $0 as compared to a cost of $100, this will dramatically inflate their individual threshold for cost. Conversely, the TT has bounds in its values—an individual can only have a cost threshold between $0 and $3200—as the items only ask about these specific levels. This can be seen in more detail when viewing the distributions of the DCE (at the level of the individual) and TT thresholds in Figure 3. Here, the spikes that can occur within the TT method are visible. As there are only discrete values that the TT threshold can take, the distribution will be much more peaked as compared to the individual-level DCE thresholds that are calculated using ratios and thus can be considered more continuous.

|  |
| --- |
|  |
| Figure 3. Distributions for TT and DCE methods.  Note that all these tables are cut down to include only those people that existed within the real range for false positive/negative (0-100), and within the matching range for threshold technique for cost (0-3500). |

Relationships between DCE, TT, and ESCAPE*.* As can be seen in Figure 4, threshold values for DCE and TT are largely uncorrelated. The largest correlation present is a small .23 correlation between the DCE false negative threshold and the TT false negative threshold. While TT values are all slightly correlated and many ESCAPE subscales are slightly correlated, no DCE values are correlated with any other DCE values. The most impressive correlations found here are between the accuracy and benefit subscales of ESCAPE (*r*=.52, indicating that those who believed accuracy of a test was important also believed the benefit of the test was important), and the ESCAPE cost subscale and the TT cost threshold (*r*=-.35, indicating that those who reported cost as being a more important consideration were more likely to have lower cost thresholds).

|  |
| --- |
|  |
| Figure 4. Correlations between DCE and TT thresholds and ESCAPE subscales |

Predicting Threshold and ESCAPE Values*.* Given a variety of information regarding the background and demographics of our sample, we wanted to understand if any of these variables could predict our threshold values, as well as the ESCAPE subscales. As can be seen in Table 12 there is no clear pattern in predicting these values. For instance, while being scared of getting cancer is predictive of the benefit and familiarity scores on the subscales of ESCAPE, they are not important variables when predicting thresholds.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 12. Estimates and standard errors when predicting thresholds and subscales with demographic data. | | | | | | | | | | | |
|  | DCE | | | TT | | | ESCAPE | | | | |
| Predictor | False Pos. | False Neg. | Cost | False Pos. | False Neg. | Cost | Accuracy | Benefit | Familiarity | Cost | Fit |
| Intercept | -47257.54 (39220.07) | 261.01 (655.07) | -9.90E+13 (8.58E+13) | -5.28 (22.81) | 3.92 (7.92) | 328.11 (1025.04) | **6.44 (0.61) \*\*\*** | **4.69 (0.71) \*\*\*** | 1.77 (1.02) | **4.97 (1.45) \*\*\*** | 2.24 (1.48) |
| Age | 3089.98 (1738.11) | -10.75 (29.03) | 5.02E+12 (3.80E+12) | 1.248 (1.01) | -0.14 (0.35) | -45.17 (45.43) | -0.05  (0.03) | 0.00 (0.03) | 0.03 (0.04) | 0.07 (0.06) | 0.06 (0.07) |
| Previously Screened | -3657.32 (12590.8) | -16.95 (210.30) | -1.67E+13 (2.76E+13) | 3.66 (7.32) | 0.24 (2.54) | -77.53 (329.07) | 0.00 (0.20) | 0.04 (0.23) | -0.22 (0.33) | 0.50 (0.47) | -0.20 (0.47) |
| Want to be screened in future | 2298.26 (8182.00) | 107.67 (136.66) | 5.09E+12 (1.79E+130 | 2.50 (4.76) | 0.28 (1.65) | 213.72 (213.84) | 0.16 (0.13) | -0.01 (0.15) | -0.21 (0.21) | -0.44 (0.30) | 0.01 (0.31) |
| Worried about cancer | 160.99 (2147.00) | -27.59 (35.86) | -2.88E+11 (4.70E+12) | -2.03 (1.25) | -0.16 (0.43) | -40.78 (56.11) | 0.01 (0.03) | -0.02 (0.04) | -0.01 (0.06) | -0.13 (0.08) | -0.01 (0.08) |
| Scared of getting cancer | -3820.77 (2147.00) | -32.85 (38.29) | -6.32E+12 (5.02E+12) | 0.32 (1.33) | 0.46 (0.46) | 35.80 (59.92) | 0.05 (0.04) | **0.15 (0.04) \*\*\*** | **0.12 (0.06) \*** | -0.03 (0.08) | 0.06 (0.09) |
| Likely to get cancer | -466.89 (2347.24) | **101.62 (39.20) \*** | 2.73E+12 (5.14E+12) | **3.15 (1.37) \*** | 0.58 (0.47) | 23.33 (61.35) | -0.02 (0.04) | -0.03 (0.04) | -0.10 (0.06) | 0.08 (0.09) | -0.09 (0.09) |
| First degree relatives with cancer | -68.27 (3214.60) | 36.31 (53.69) | -7.36E+12 (7.03E+12) | 1.39 (1.87) | **1.61 (0.65) \*** | 139.57 (84.02) | 0.05 (0.05) | 0.04 (0.06) | -0.01 (0.08) | -0.08 (0.12) | -0.01 (0.12) |
| Second degree relatives with cancer | -1061.79 (1651.09) | -7.85 (27.58) | **7.87E+12 (3.61E+12) \*** | 0.09 (0.96) | -0.09 (0.33) | -50.99 (43.15) | -0.02 (0.03) | -0.04 (0.03) | -0.07 (0.04) | **-0.12 (0.06) \*** | -0.08 (0.06) |
| Friends with cancer | -240.14 (1947.69) | -25.22 (32.53) | -4.62E+12 (4.26E+12) | -1.23 (1.13) | -0.36 (0.39) | 26.38 (50.9) | -0.00 (0.03) | 0.01 (0.04) | 0.08 (0.05) | 0.09 (0.07) | 0.05 (0.07) |
| Sex | -7387.51 (6172.65) | 46.462 (103.10) | -1.27E+13 (1.35E+13) | -1.15 (3.59) | -0.45 (1.25) | -33.47 (161.33) | -0.12 (0.10) | -0.14 (0.11) | **0.32 (0.16) \*** | 0.35 (0.23) | -0.09 (0.23) |
| Race | 2916.43 (7576.86) | 111.72 (126.55) | 8.25E+12 (1.66E+13) | **-12.47 (4.41) \*\*** | 0.90 (1.53) | -258.77 (198.03) | -0.06 (0.12) | 0.18 (0.14) | -0.02 (0.20) | -0.42 (0.28) | -0.21 (0.29) |
| SNS | 1512.4 (2948.59) | -41.39 (49.25) | 9.56E+10 (6.45E+12) | -0.00 (1.71) | -0.29 (0.60) | 126.34 (77.06) | 0.08 (0.05) | 0.09 (0.05) | -0.14 (0.08) | -0.21 (0.11) | 0.07 (0.11) |
| MMS | 1988.4 (2770.21) | -3.42 (46.27) | 5.87E+12 (6.06E+12) | 1.58 (1.61) | 0.45 (0.56) | 141.14 (72.4) | 0.056 (0.04) | **0.12 (0.05) \*** | **0.34 (0.07) \*\*\*** | 0.07 (0.10) | 0.18 (0.10) |
| Note: \* indicates p <.05, \*\* indicates p<.01, \*\*\* indicates p<.001 | | | | | | | | | | | |

61

Predicting Choice with Thresholds and ESCAPE*.* Finally, we sought to identify which of these measures of preference, if any, could predict whether or not these individuals would choose to receive a cancer screening test without benefit (in terms of lives saved). As can be seen in Table 13, none of our measures of preference predicted choice in this hypothetical scenario.[[5]](#footnote-5)

|  |  |  |
| --- | --- | --- |
| Table 13. Model estimates when predicting choice with DCE thresholds, TT thresholds, and ESCAPE subscales | | |
|  | B (SE) | *p* |
| Intercept | 0.52 (1.32) | 0.70 |
| DCE False Positive | 0.00 (0.00) | 0.24 |
| DCE False Negative | 0.00 (0.00) | 0.81 |
| DCE Cost | 0.01 (0.00 | 0.99 |
| TT False Positive | 0.00 (0.01) | 0.61 |
| TT False Negative | 0.01 (0.02) | 0.45 |
| TT Cost | -0.00 (0.00) | 0.91 |
| Accuracy | -0.10 (0.22) | 0.63 |
| Benefit | 0.05 (0.19) | 0.77 |
| Familiarity | 0.18 (0.12) | 0.14 |
| Cost | -0.06 (0.09) | 0.49 |
| Fit | -0.12 (0.09) | 0.16 |

*Conclusions and Limitations*

We sought to compare the ESCAPE scale to thresholds from both the DCE and TT methods. While sample level DCE values were similar to overall TT values, this was not true when considering individual level DCE values because the method of computation creates extremely skewed ratios for many individuals. Although intended to measure similar ideas, DCE thresholds, TT thresholds, and ESCAPE subscales were not found to be highly correlated. Although false negative thresholds for DCE and TT were slightly correlated and the TT cost threshold was slightly correlated with the ESCAPE cost subscale, no other noteworthy correlations existed between the three methods.

When attempting to understand what may drive these threshold and subscale responses, we found no consistent predictors. Further, when attempting to predict choice, none of our thresholds or subscales were significant predictors. These findings are somewhat disheartening as we would have hoped at least that our three methods of measuring preference were measuring similar things, and that our threshold and subscale values could predict choice.

Limitations of this study include the sample of undergraduate students, who may not have sufficient exposure to screening tests to respond in a consistent manner and who may have not given the tasks sufficient attention. However, as these results did not appreciably change when low quality data individuals were removed from the sample, it is unlikely that lack of attention caused these results.

**Chapter 9**

**Discussion and Future Directions**

This dissertation worked toward the development of a scale focusing on individuals’ preferences for attributes of screening tests. To do this, we created a set of possible screening test attributes from relevant research articles (Chapter 2), ran an exploratory factor analysis to identify the underlying structure of these items (Chapter 3), and completed a confirmatory factor analysis to confirm the structure of the items (Chapter 4). In doing so, we identified 5-Factors that our undergraduates deemed extremely important: the accuracy of the test, the possible benefits that could arise from being tested, how familiar the test was, the monetary cost of the test, and how it fit with their personal preferences.

Beyond identifying a factor structure, we showed that the scale was relatively reliable (Chapter 5) Once outlying observations were removed from our dataset, we found that the benefit, familiarity, and personal preference subscales could reach acceptable reliability values, but that the accuracy and cost subscales were less robust.

After this, the convergent and discriminant validity of the scale was assessed (Chapter 6). Here we found that importance of accuracy was positively related to variables such as being scared of cancer, working well with numbers, and being more aware of your health, and negatively related to being more risk taking across domains. Believing the benefit of a test was important was positively related to being worried or scared about cancer and being a medical maximizer, and it was negatively related to being risk taking in ethical situations. The importance of cost was positively related to being worried about paying for an unexpected medical expense, avoiding the doctor due to cost, and getting more insurance if you could afford it. The importance of personal preference was positively related to being more averse to ambiguity, having more faith and meaning in life, having a more internal or external religious orientation, and being more religious. Interestingly, the importance of familiarity was not related to any of the included scales or items.

Further, while investigating whether or not the structure could be replicated in a more diverse population (Chapter 7), we demonstrated that the ESCAPE items have different underlying factor structures in the undergraduate student sample when compared to a more diverse sample. While the accuracy and benefit factors were similar between the two cohorts, new factors arose as well. These new factors included the ease/comfort of test, a need for action, and disease attributes. Additionally, while the familiarity factor consisted of three items in the more diverse sample, it contained only two of these items in the undergraduate sample (e.g. “encouraged by someone on TV” and “family member or friend has had before”). It could be argued that while these items are slightly different, they hold a similar message.

Finally, we compared the ESCAPE scale with other methods of measuring preferences (Chapter 8), including a Discrete Choice Experiment (DCE) and a Threshold Technique (TT). We found similarities between the TT and DCE threshold values when the DCE thresholds were calculated at the level of the sample. However, these estimates became extreme at the individual level and thus TT differed from DCE thresholds at the individual level. Importantly, although all three of these methods intended to measure similar preferences, correlations between DCE thresholds (at the individual level), TT thresholds, and ESCAPE subscales were quite small. While false negative thresholds for DCE and TT were slightly correlated and the TT cost threshold was slightly correlated with the ESCAPE cost subscale, no other noteworthy correlations existed between the three methods. This point and the previous point are critical as each of these measures purportedly informs researchers of patient preferences. If all methods are thought to measure preferences in the same way and the reality is that each method offers varying estimates of preference (as this work suggests) then it is vital that we identify what method of preference elicitation is used and qualify these results. Research on preferences may be skewed or inaccurate if we were to believe that all elicitation technique are equal and provide the same information. Further, when attempting to understand what may impact these thresholds and subscale responses, we found no consistent predictors. This means that there were no demographic or other variables that were capable of predicting what was important to these individuals in any consistent manner.

Most importantly, when attempting to predict whether or not an individual would choose to be screened by a cancer screening test that held no benefit in terms of lives saved, none of our thresholds or subscales were significant predictors of choice. This indicated that not only do these three methods measure preferences in three categorically different ways, but also that none of these varying measures of preferences were capable of capturing information that would be uniquely predictive of choice. While this is only a single study, these results suggest that all three of these methods may not be useful at identifying how these types of screening preferences impact decisions. Although these studies routinely speak to population-level preferences, they rarely use these preferences to predict choices. As current health policies are informed by these sorts of preference elicitation methods these policies may be inaccurate or, at the very least, imprecise.

Future work on the ESCAPE scale first aims to conduct a Confirmatory Factor Analysis on a sample from a move diverse population to be able to confirm the structure identified in the previous Exploratory Factor Analysis presented here in Chapter 7. If a factor structure can be confirmed in this more diverse population, this work should additionally seek to identify if the scale is reliable and valid, and if it can predict screening intentions.

Additionally, the current data at hand could be re-analyzed using an exploratory structural equation model which could allow us to compare the undergraduate and Mturk samples by setting certain items to load on designated factors while allowing other items to load freely where they may best fit. This could further our understanding of the similarities and differences between these two samples. Further, as such striking differences were found between undergraduates and a more diverse population, it would be wise to test the factor structure of these items specifically within an older population (aged 40-70) of individuals that will be most likely to be making these kinds of decisions such as whether to be screened, when to start screening, and how frequently to screening, and among an even older population of adults (aged 70-85) where screening cessation decisions need to be made.

As this age difference may be a key point of investigation, a future study should aim to identify the trajectories of the development of these preferences over time. If these preferences are repeatedly measured alongside pertinent life events and healthcare experiences, we may glean a better understanding of what is creating or altering these preferences, as well as how predictive these preferences may be of future preferences and choices about screening.

Finally, one future goal of this scale is to tailor messages to individuals based upon their responses to this scale. As such, additional work should attempt to identify how people high or low on the various subscale factors respond to different messages regarding the risks and benefits of screening. For instance, an individual who is low on the familiarity factor but high on the ease/discomfort factor may respond better to information that focuses on the newest, less invasive forms of screening for colon cancer, whereas an individual who is high on the accuracy subscale may be more interested in learning about more conventional, invasive colon cancer screening techniques. Further, it is quite possible that one subscale outweighs all other subscales—perhaps for both these individuals the decision only comes down to the benefit related to each type of screening test. Using the factors we know are important to that population, we should identify messages that efficiently explain the risks and benefits to help increase coherence.

Toward this goal, we should additionally look into the best ways to present this information. As the attitudes around health care are evolving, it is possible that discussions surrounding health and healthcare are now increasingly more like political discussions—individuals usually enter into the discussion with distinct values and preferences and attempts to alter these may take more than recommendations or data-based presentations of information. In the increasingly complex world of medical decisions, we must now take into account things such as trust in the physician, trust in healthcare, and alignment between the beliefs of the recommending physician and the patient.

Within this dissertation I have created a set of factors regarding screening test attributes and both explored and confirmed a 5-factor structure. I have also shown that the scale is reliable and has convergent and discriminant validity. Further, I have identified that this structure cannot be replicated in a more diverse population, and instead that the more diverse sample has a 6-factor structure. Finally, I have compare this individual difference scale with a DCE and a TT, finding all of these methods vary and none of them are capable of predicting screening choices.

Overall, we find that while some preferences are widely shared (the importance of a test being accurate), other preferences vary (the importance of cost), or do not exist within particular populations (importance of taking action). Thus, when attempting to tailor information to the individual we must understand the features that a person cares about—and be aware that these may fluctuate with age and differ by ethnicity or other factors—in order to properly address their needs. Additionally, we must be aware that different preference elicitation techniques can produce drastically different measures of preferences. Further, these measures of what people say they prefer may not actually map onto their actions. Therefore, we need to identify whether choices that are inconsistent with stated preferences arise due to the way preferences were measured or due to a true inconsistent behavior of the individual.

References

American Cancer Society, . (2017a). Cancer Screening Guidelines By Age. Retrieved January 1, 2017, from https://www.cancer.org/healthy/find-cancer-early/cancer-screening-guidelines/screening-recommendations-by-age.html

American Cancer Society, . (2017b). Colorectal Cancer Screening Tests. Retrieved January 1, 2018, from https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/screening-tests-used.html

Bibbins-Domingo, K., Grossman, D. C., & Curry, S. J. (2017). The US Preventive Services Task Force 2017 Draft Recommendation Statement on Screening for Prostate Cancer: An Invitation to Review and Comment. *JAMA: The Journal of the American Medical Association*. https://doi.org/https://doi.org/10.1001/jama.2017.4413%0D

Blais, A.-R., & Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, *1*(1), 33–47.

Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-1035-6

Buchanan, E. M., Valentine, K. D., & Schulenberg, S. E. (2014). Exploratory and confirmatory factor analysis: Developing the Purpose in Life test –Short Form (PIL-SF). In P. Bindle (Ed.), *Sage Research Methods Cases* (pp. 1–18). Sage. https://doi.org/10.4135/978144627305013517794

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* Routledge.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10*(7), 1–9.

Davis, T. C., Arnold, C., Berkel, H., & Nandy, I. (1996). Knowledge and attitude on screening mammography among low‐literate, low‐income women. *Cancer*, 1912–1920. https://doi.org/10.1002/(SICI)1097-0142(19961101)78:9<1912::AID-CNCR11>3.0.CO;2-0 [pii]

DeAngelis, C. D., & Fontanarosa, P. B. (2010). US Preventive Services Task Force and breast cancer screening. *JAMA : The Journal of the American Medical Association*, *303*(2), 172–3. https://doi.org/10.1001/jama.2009.1990

Dolan, J. G., & Frisina, S. (2002). Randomized Controlled Trial of a Patient Decision Aid for Colorectal Cancer Screening. *Medical Decision Making*, *22*(2), 125–139. https://doi.org/10.1177/0272989X0202200210

Dolan, N. C., Ferreira, M. R., Davis, T. C., Fitzgibbon, M. L., Rademaker, A., Liu, D., … Bennett, C. L. (2004). Colorectal cancer screening knowledge, attitudes, and beliefs among veterans: Does literacy make a difference? *Journal of Clinical Oncology*, *22*(13), 2617–2622. https://doi.org/10.1200/JCO.2004.10.149

Ehrenreich, B. (2009). The uproar over new breast cancer screening guidelines. *The Huffıngton Post*.

Esserman, L. J., Thompson, I. M., & Reid, B. (2013). Re: Overdiagnosis and overtreatment in cancer: An opportunity for improvement. *JAMA*, *310*, 797–798. https://doi.org/10.1016/j.eururo.2013.10.023

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*, *27*(5), 672–680. https://doi.org/10.1177/0272989X07304449

Feinleib, M., & Zelen, M. (1969). Some pitfalls in the evaluation of screening programs. *Arch Environ Health*, *19*, 412–415.

Feldman-stewart, D., Tong, C., Siemens, R., Alibhai, S., Pickles, T., Robinson, J., & Brundage, M. D. (2012). The Impact of Explicit Values Clarification Exercises in a Patient Decision Aid Emerges After the Decision Is Actually Made : Evidence From a Randomized Controlled Trial, 616–626. https://doi.org/10.1177/0272989X11434601

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

Goodwin, J. S., Tan, A., Jaramillo, E., & Kuo, Y. F. (2013). Prostate-specific antigen testing in men aged 40-64 years: Impact of publication of clinical trials. *Journal of the National Cancer Institute*, *105*(10), 743–745. https://doi.org/10.1093/jnci/djt039

Gorsuch, R., & Venable, G. (1983). Development of an" Age Universal" IE scale. *Journal for the Scientific Study of Religion*, *22*(2), 181–187. https://doi.org/10.2307/1385677

Greiner, K. A., Born, W., Nollen, N., & Ahluwalia, J. S. (2005). Knowledge and perceptions of colorectal cancer screening among urban African Americans. *Journal of General Internal Medicine*, *20*(11), 977–983. https://doi.org/10.1111/j.1525-1497.2005.0165.x

Grossman, D. C., Curry, S. J., Owens, D. K., Bibbins-Domingo, K., Caughey, A. B., Davidson, K. W., … Tseng, C. W. (2018). Screening for prostate cancer USPreventive servicestaskforcerecommendation statement. *JAMA - Journal of the American Medical Association*, *319*(18), 1901–1913. https://doi.org/10.1001/jama.2018.3710

Guttmann, R., Castle, R., & Fiebig, D. G. (2009). Use of Discrete Choice Experiments in health economics: An update of the literature. *Health Economics*, (May), 23. https://doi.org/10.1016/j.techfore.2015.12.014

Gyrd-Hansen, D., & Søgaard, J. (2001). Analysing Public Preferences for Screening Programmes. *Health Economics*, *10*, 617–634. https://doi.org/10.1002/hec.622

Han, P. K., Reeve, B. B., Moser, R. P., & Klein, W. M. (2009). Aversion to ambiguity regarding medical tests and treatments: Measurement, prevalence, and relationship to sociodemographic factors. *Journal of Health Communication*, *14*(6), 556–572. https://doi.org/doi:10.1080/10810730903089630

Hawley, S. T., McQueen, A., Bartholomew, L. K., Greisinger, A. J., Coan, S. P., Myers, R., & Vernon, S. W. (2012). Preferences for colorectal cancer screening tests and screening test use in a large multispecialty primary care practice. *Cancer*, *118*(10), 2726–2734. https://doi.org/10.1002/cncr.26551

Hawley, S. T., Zikmund-Fisher, B., Ubel, P., Jancovic, A., Lucas, T., & Fagerlin, A. (2008). The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Education and Counseling*, *73*(3), 448–455. https://doi.org/10.1016/j.pec.2008.07.023

Herbst, R. S., Bajorin, D. F., Bleiberg, H., Blum, D., Hao, D., Johnson, B. E., … Bentsen, T. (2006). Clinical Cancer Advances 2005: Major research advances in cancer treatment, prevention, and screening - A report from the American Society of Clinical Oncology. *Journal of Clinical Oncology*, *24*(1), 190–205. https://doi.org/10.1200/JCO.2005.04.8678

Hoffman, R. M., Elmore, J. G., Pignone, M. P., Gerstein, B. S., Levin, C. A., & Fairfield, K. M. (2016). Knowledge and values for cancer screening decisions: Results from a national survey. *Patient Education and Counseling*, *99*(4), 624–630. https://doi.org/https://doi.org/10.1016/j.pec.2015.11.001

Horne, R., Weinman, J., & Hankins, M. (1999). The beliefs about medicines questionnaire: The development and evaluation of a new method for assessing the cognitive representation of medication. *Psychology & Health*, *14*, 1–24. https://doi.org/http://dx.doi.org/10.1080/08870449908407311

Howard, D. H., & Adams, E. K. (2012). Mammography rates after the 2009 US Preventive Services Task Force breast cancer screening recommendation. *Preventive Medicine,* *55*(5), 485–487.

Imaeda, A., Bender, D., & Fraenkel, L. (2010). What is most important to patients when deciding about colorectal screening? *Journal of General Internal Medicine*, *25*(7), 688–693. https://doi.org/10.1007/s11606-010-1318-9

Kamerow, D., & Woolf, S. (2009). Parsing the mammogram guidelines. *The Washington Post*, p. A25.

Kang, S. K., Scherer, L. D., Megibow, A. J., Higuita, L. J., N., K., Braithwaite, R. S., & Fagerlin, A. (2018). A randomized study of patient risk perception for incidental renal findings on diagnostic imaging tests. *American Journal of Roentgenology*, *210*, 369–375. https://doi.org/10.2214/AJR.17.18485

Kattan, M. W. (2016). The Hypothetical Rabbit. *Frontiers in Oncology*, *6*(May), 1–2. https://doi.org/10.3389/fonc.2016.00123

Kennedy, E. D., Schmocker, S., Victor, C., Baxter, N. N., Kim, J., Brierley, J., & McLeod, R. S. (2011). Do patients consider preoperative chemoradiation for primary rectal cancer worthwhile? *Cancer*, *117*(13), 2853–2862. https://doi.org/10.1002/cncr.25842

Koopmans, T. C. (1964). On flexibility of future preference. In S. Maynard & B. Glenn (Eds.), *Human Judgements and Optimality* (pp. 243–254). New York: Wiley.

Kopec, J. A., Richardson, C. G., Llewellyn-thomas, H., Klinkhoff, A., Carswell, A., & Chalmers, A. (2017). Probabilistic threshold technique showed that patients ’ preferences for specific trade-offs between pain relief and each side effect of treatment in osteoarthritis varied, *60*(2007). https://doi.org/10.1016/j.jclinepi.2007.01.001

Krantz, D. S., Baum, A., & von Wideman, M. (1980). Assesment of preferences for self-treatment and information in health care. *Journal of Personality and Social Psychology*, *39*(5), 977–990.

Laine, C., Dickersin, K., & Mulrow, C. (2016). Time to Douse the Firestorm Around Breast Cancer ScreeningTime to Douse the Firestorm Around Breast Cancer Screening. *Annals of Internal Medicine*, *164*(4), 303–304. https://doi.org/10.7326/M15

Lau, R. R., Hartman, K. A., & Ware, J. E. (1986). Health as a value: Methodological and theoretical considerations: EBSCOhost, *5*(1), 25–43. Retrieved from http://web.a.ebscohost.com.ezproxy.bu.edu/ehost/detail/detail?sid=b7b16a6f-3eb2-42f3-a42d-c1c6e65fd25e%40sessionmgr4004&vid=1&hid=4109&bdata=JnNpdGU9ZWhvc3QtbGl2ZSZzY29wZT1zaXRl#db=pdh&AN=1987-21566-001&anchor=c36

Ledesma, R. D., & Valecro-Mora. (2007). Determining the number of factors to retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis. *Practical Assessment, Research & Evaluation*, *12*(2), 2–11. https://doi.org/http://pareonline.net/getvn.asp?v=12&n=2

Ling, B. S., Moskowitz, M. A., Wachs, D., Pearson, B., & Schroy, P. C. (2001). Attitudes towards colorectal cancer screening tests: a survey of patients and physicians. *J Gen Intern Med*, *16*(12), 822–830.

Longley, S. L., Watson, D., & Noyes, R. (2005). Assessment of the hypochondriasis domain: The Multidimensional Inventory of Hypochondriacal Traits (MIHT). *Psychological Assessment*, *17*(1), 3–14. https://doi.org/10.1037/1040-3590.17.1.3

Mansfield, C., Tangka, F. K., Ekwueme, D. U., Smith, J. L., Guy Jr, G. P., Li, C., & Hauber, A. B. (2016). Peer Reviewed: Stated Preference for Cancer Screening: A Systematic Review of the Literature, 1990–2013. *Preventing Chronic Disease*, *13*.

Moyer, V. a. (2012). Screening for prostate cancer: U.S. preventive services task force recommendation statement. *Ann Intern Med*, *157*(2), 120–134. https://doi.org/10.7326/0003-4819-157-2-201207170-00459

Mulhem, E., Fulbright, N., & Duncan, N. (2015). Prostate cancer screening. *American Family Physician*, *92*(8), 683–688. https://doi.org/10.1016/S0002-838X(15)30301-4

Muthén, L. K., & Muthén, B. O. (2010). *Mplus Statistical Analysis With Latent Variables USER ’ s Guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Nayaradou, M., Berchi, C., Dejardin, O., & Launoy, G. (2010). Eliciting population preferences for mass colorectal cancer screening organization. *Medical Decision Making*, *30*(2), 224–233. https://doi.org/10.1177/0272989X09342747

Nickel, B., Barratt, A., Hersch, J., Moynihan, R., Irwig, L., & McCaffery, K. (2015). How different terminology for ductal carcinoma in situ (DCIS) impacts women’s concern and management preferences: A qualitative study. *The Breast*, *24*(5), 673–679.

O’Connor, A. M., Rostom, A., Fiset, V., Tetroe, J., Entwistle, V., Llewellyn-Thomas, H., … Jones, J. (1999). Decision aids for patients facing health treatment or screening decisions: systematic review. *Bmj*, *319*(7212), 731–734. https://doi.org/10.1136/bmj.319.7212.731

Oeffinger, K. C., Fontham, E. T., Etzioni, R., Herzig, A., Michaelson, J. S., Shih, Y.-C. T., & Wender, R. (2015). Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. *JAMA*, *314*(15), 1599. https://doi.org/https://doi.org/10.1001/jama.2015.12783

Omer, Z. B., Hwang, E. S., Esserman, L. J., Howe, R., & Ozanne, E. M. (2013). Impact of ductal carcinoma in situ terminology on patient treatment preferences. *JAMA Internal Medicine*, *173*(19), 1830–1831.

Palmer, R. C., Midgette, L. A., & Mullan, I. D. (2010). Colorectal cancer screening preferences among African Americans: Which screening test is preferred? *Journal of Cancer Education*, *25*(4), 577–581. https://doi.org/10.1007/s13187-010-0081-2

Paulos, J. A. (1988). *Innumeracy: Mathematical illiteracy and its consequences*. Macmillan.

Perez, T. Y., Danzig, M. R., Ghandour, R. A., Badani, K. K., Benson, M. C., & McKiernan, J. M. (2015). Impact of the 2012 United States Preventive Services Task Force statement on prostate-specific antigen screening: analysis of urologic and primary care practices. *Urology*, *85*(1), 85–91.

Peterman, A. H., Fitchett, G., Brady, M. J., Hernandez, L., & Cella, D. (2002). Measuring spiritual well-being in people with cancer: The functional assessment of chronic illness therapy—spiritual well-being scale (FACIT-Sp). *Annals of Behavioral Medicine*. https://doi.org/10.1207/S15324796ABM2401\_06

Phillips, K. A., Van Bebber, S., Marshall, D., Walsh, J., & Thabane, L. (2006). A review of studies examining stated preferences for cancer screening. *Preventing Chronic Disease*, *3*(3), A75–A75. https://doi.org/A75 [pii]

Pleis, J. R., Lucas, J. W., & Ward, B. W. (2009). Summary health statistics for U.S. adults: National Health Interview Survey, 2008. *Vital and Health Statistics.*, *Series 10,*(242), 1–157.

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria.

Rai, T., Clements, A., Bukach, C., Shine, B., Austoker, J., & Watson, E. (2007). What influences men’s decision to have a prostate-specific antigen test? A qualitative study. *Family Practice*, *24*(4), 365–371. https://doi.org/10.1093/fampra/cmm033

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Revelle, W. (2018). psych: Procedures for Personality and Psychological Research. Evanston, Illinois, USA: Northwestern University. Retrieved from https://cran.r-project.org/package=psych

Revelle, W., & Rocklin, T. (1979). Very simple structure – alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, *14*(4), 403–414.

Rice, C. E., Dandreaux, D., Handley, E. D., & Chassin, L. (2006). Children of Alcoholics: Risk and Resilience. *Prevention Researcher*, *13*(4), 3–6. Retrieved from http://ezproxy.csp.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ793223%5Cnhttp://www.tpronline.org/product\_toc/13-4-toc.pdf

Rosenberg, R., Yankaskas, B., Abraham, L., Sickles, E., Lehman, C., Geller, B., … Ballard-Barbash, R. (2006). Performance benchmarks for screening mammography. *Radiology*, *241*, 55–66.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software,* *48*(2), 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/

Rostosky, S. S., Regnerus, M. D., & Wright, M. L. C. (2017). Coital Debut : The Role of Religiosity and Sex Attitudes in the Add Health Survey. *The Journal of Sex Research*, *40*(4), 358–367.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, *80*(1), 1–28.

Ryan, M. (2004). Discrete choice experiments in health care. *Bmj*, *328*(7436), 360–361. https://doi.org/10.1136/bmj.328.7436.360

Ryan, M., & Gerard, K. (2003). Using Discrete Choice Experiments to Value Health Care Programmes: Current Proactice and Future Research Reflections. *Applied Health Economics and Health Policy*, *2*(1), 55–64.

Scaf-Klomp, W., Sanderman, R., van de Wiel, H. B., Otter, R., & van den Heuvel, W. J. (1997). Distressed or relieved? Psychological side effects of breast cancer screening in The Netherlands. *Journal of Epidemiology & Community Health*, *51*(6), 705–710. https://doi.org/10.1136/jech.51.6.705

Scherer, L. D., Caverly, T. J., Burke, J., Zikmund-Fisher, B. J., Kullgren, J. T., Steinley, D., … Fagerlin, A. (2016). Development of the medical maximizer-minimizer scale. *Health Psychology*, *35*(11), 1276–1287. https://doi.org/10.1037/hea0000417

Scherer, L. D., Valentine, K. D., Patel, N., Baker, S. G., & Fagerlin, A. (2018). A Bias for Action in Cancer Screening? *(Unpublished Manuscript)*.

Schroy, P. C., Emmons, K., Peters, E., Glick, J. T., Robinson, P. A., Lydotes, M. A., … Pignone, M. (2011). The impact of a novel coputer-based decision aid on shared decision-making for colorectal cancer screening: A randomized trial. *Medical Decision Making*, *31*(1), 93–107. https://doi.org/10.1177/0272989X10369007.The

Schwartz, L. M., Woloshin, S., & Fowler Jr, F. J. (2004). Enthusiasm for cancer screening in the United States. *JAMA : The Journal of the American Medical Association*, *291*(1), 71–78.

Shokar, N. K., Vernon, S. W., & Weller, S. C. (2005). Cancer and colorectal cancer: Knowledge, beliefs, and screening preferences of a diverse patient population. *Family Medicine*, *37*(5), 341–347.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability.1. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–8. https://doi.org/10.1037/0033-2909.86.2.420

Shwartz, M. (1980). Estimates of Lead Time and Length Bias in a Breast Cancer Screening Program, 844–851.

Slovic, P. (1995). The construction of preference. *American Psychologist*, *50*, 364–371.

Squiers, L. B., Bann, C. M., Dolina, S. E., Tzeng, J., McCormack, L., & Kamerow, D. (2013). Prostate-specific antigen testing: men’s responses to 2012 recommendation against screening. *American Journal of Preventive Medicine*, *45*(2), 182–189.

Squiers, L. B., Holden, D. J., Dolina, S. E., Kim, A. E., Bann, C. M., & Renaud, J. M. (2011). The public’s response to the U.S. preventive services task force’s 2009 recommendations on mammography screening. *American Journal of Preventive Medicine*, *40*(5), 497–504. https://doi.org/10.1016/j.amepre.2010.12.027

Stafinski, T., Menon, D., Nardelli, A., Bakal, J., Ezekowitz, J., Tymchak, W., … Armstrong, P. (2015). Incorporating patient preferences into clinical trial design: Results of the Opinions of Patients on Treatment Implications of New Studies (OPTIONS) project. *Am Heart J*, *169*, 122–131.

Steele, C. B., Miller, D. S., Maylahn, C., Uhler, R. J., & Baker, C. T. (2000). Knowledge, attitudes, and screening practices among older men regarding prostate cancer. *American Journal of Public Health*, *90*(10), 1595–1600. https://doi.org/10.2105/AJPH.90.10.1595

Swets, J. A., & Swets, J. A. (1988). Measuring the Accuracy of Diagnostic Systems Linked references are available on JSTOR for this article : Measuring the Accuracy of Diagnostic Systems. *Science*, *240*(4857), 1285–1293.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics.* Allyn & Bacon/Pearson Education.

Trauzettel-Klosinski, S., & Dietz, K. (2012). Standardized assessment of reading performance: The new international reading speed texts IReST. *Investigative Ophthalmology and Visual Science*, *53*(9), 5452–5461. https://doi.org/10.1167/iovs.11-8284

Trumbo, C. W., McComas, K. A., & Kannaovakun, P. (2007). Cancer anxiety and the perception of risk in alarmed communities. *Risk Analysis*, *27*(2), 337–350. https://doi.org/10.1111/j.1539-6924.2007.00886.x

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies (Personnel Research Section Report, No. 984).* Washington, D.C.

Tversky, A., & Thaler, R. H. (1990). Anomalies: preference reversals. *The Journal of Economic Perspectives*, *4*(2), 201–211.

U.S. Preventive Services Task Force, . (2016). Final Update Summary: Breast Cancer: Screening. Retrieved January 1, 2017, from https://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/breast-cancer-screening

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–7.

Victory, J. (2017). The harms of cancer screening: It’s time to get personal. Retrieved January 1, 2018, from https://www.healthnewsreview.org/2017/02/the-harms-of-cancer-screening-its-time-to-get-personal/

Waller, J., Osborne, K., & Wardle, J. (2015). Enthusiasm for cancer screening in Great Britain: A general population survey. *British Journal of Cancer*, *112*(3), 562–566. https://doi.org/10.1038/bjc.2014.643

Wallston, B., Wallston, K. A., Kaplan, G. D., & Maides, S. A. (1976). Development and validation of the Health Locus of Control scale. *J Consult Clin Psychol*, *44*(4), 580–585.

Welch, H. G., & Black, W. C. (2010). Overdiagnosis in cancer. *Journal of the National Cancer Institute*, *102*(9), 605–613. https://doi.org/10.1093/jnci/djq099

Wilt, T., Paul, J., Murdoch, M., Nelson, D., Nugent, S., & Rubins, H. (2001). Educating men about prostate cancer screening. A randomized trial of a mailed pamphlet. *Eff Clin Pract*, *4*, 112–20.

Wolf, A. M., Wender, R. C., Etzioni, R. B., Thompson, I. M., Amico, A. Vd, Volk, R. J., … Smith, R. a. (2010). American Cancer Society Guideline for the Early Detection of Prostate Cancer Update 2010. *Cancer Journal, The*, *60*(2), 70–98. https://doi.org/10.3322/caac.20066.Available

Wolf, R. L., Basch, C. E., Brouse, C. H., Shmukler, C., & Shea, S. (2006). Patient preferences and adherence to colorectal cancer screening in an urban population. *American Journal of Public Health*, *96*(5), 809–811. https://doi.org/10.2105/AJPH.2004.049684

Woloshin, S., & Schwartz, L. M. (2010). The benefits and harms of mammography screening: Understanding the trade-offs. *JAMA - Journal of the American Medical Association*, *303*(2), 164–165. https://doi.org/10.1001/jama.2009.2007

**Appendix I**

How-tos

1) Download box sync and make sure the stuff in your box.com is showing up properly on your desktop.

**2) Remember that the whole reason behind this is to find out what it is that people care about when it comes to screening test, what makes them tick. So anything discussing what predicted wanting a screening test, or how people felt about screening tests is what we want to focus on. Any scales or questions already asked similar to the questionnaire we are trying to develop should definitely be marked as such. Anything that can help make sure that we have all the important questions to ask people should be marked. Air on the side of over-inclusion, I can always go through and note that something wasn’t what I’m interested in.**

3) Go to scholar.google.com and enter in the search bar the following:

“screening test” preferences

4) Find the last mentioned article in the LitReview.xlsx spreadsheet, and continue on to the next article on the Google list.

5) Spreadsheet titles refer to the following:

Cited by: number of times cited as given by Google scholar

Link to article: the link to the full article. If full article cannot be accessed, first type

“ABSTRACT ONLY” into cell, then paste in the link where the abstract is located.

Year, Title, Authors, Journal all refer to those elements of the article

Type of article: what the focus of the article is, for instance, if it’s reporting on old data it

Would be a retrospective study, if it’s reporting on survey data, state that, if there’s been some sort of randomized control trial (RCT) preformed, label it as that. Also indicate if the population is of patients, physicians, college students, etc. If you’re not sure what an article qualifies as, email me and we can discuss.

Screening for what: could be cancers, prenatal, etc. separate out individual cancers (i.e.

colon, breast, liver)

Country of population: where was the study conducted

IVs: what was varied in the study

DVs: what were the most important outcomes measured in the study

Other measures: anything else it is clear they collected data on

Importance: state here what we may be able to pull out of the article for the purposes of

rationalizing or creating the questionnaire. If nothing is applicable, that’s ok, just leave it blank. When in doubt, include something here, even if it’s a bit tangential.

6) We will each do 10 articles a week (we can change this number depending on how this goes through the semester). My aim is to go through the top 100 (at least) of these articles.

When in doubt, send me an email about any questions.

**Appendix II**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Demographics for all studies | | | | | | | |
|  |  | Ch 3 | Ch 4 | Ch5 | Ch 6 | Ch 7 | Ch 8 |
| Age | M (SD) | 19.16 (1.01) | 18.59 (1.50) | 19.07 (0.97) | 18.81 (1.45) | 37.96 (11.70) | 18.52 (1.45) |
|  | Range | 18-25 | 17-43 | 18-22 | 18-33 | 19-72 | 18-38 |
| Gender | Female | 49.87% | 62.55% | 61.31% | 61.70% | 48.67% | 75.52% |
| Race | Caucasian | 79.20% | 80.31% | 83.21% | 81.23% | 79.67% | 87.59% |
|  | African-American | 9.07% | 6.74% | 7.30% | 9.00% | 8.67% | 4.48% |
|  | Asian/Pacific Islander | 5.33% | 2.69% | 2.19% | 4.89% | 5.33% | 2.42% |
|  | American Indian/Alaska Native | 0.53% | 0.21% | 0.73% | 0.26% | 4.00% | 0.34% |
|  | Hispanic/Latino(a) | 3.20% | 3.42% | 4.38% | 2.31% | 0.33% | 3.45% |
|  | Other/Mixed | 2.67% | 6.63 | 2.19% | 2.31% | 2.00% | 1.72% |

**Appendix III**

Original list of items organized by theoretical category

Test Accuracy (Accuracy)

Gives me accurate information about whether or not I have the disease

Will tell me that I have the disease if I do, in fact, have the disease

Will tell me that I do not have the disease if I do not, in fact, have the disease

Will give me consistent results (i.e. if I have the disease and am tested for it twice, both tests will tell me that I have the disease)

Distinguishes between diseased and non-diseased people

Is accurate at identifying people who truly have the disease

Correctly classifies people who truly do NOT have the disease

Will accurately determine the severity of the disease if it is detected

Will not lead to unnecessary treatment

Will not tell me that I have the disease when I actually do not have the disease

Will not falsely diagnose me as having the disease if I do not have the disease

Test Safety (Discomfort, complications)

Has no risk of pain

Does not have side-effects

Has no risk of anxiety

Will reassure me that I am healthy

Will not make me worry that I am unhealthy

Will not make me worried unnecessarily

Will result in minimal discomfort

Will result in only minor side effects

Will not cause serious problems

Will not lead to complications with the procedure

Is safe

Is a simple procedure

Will be a comfortable experience

Does not seem scary

Will not lead to psychological distress

Health Benefits (flipside of comparative risk/risk)

Will reduce my chances of dying from the disease

Will improve my health

Will catch the disease early

Makes me feel safer

Will help other people

Will ensure I won’t be surprised by a cancer diagnosis later in life

Will protect me from life-threatening disease

Has been shown to lead to improved health outcomes

Will find the disease early enough that I can be treated

Will increase my life expectancy

Will save my life

Will help me to live a longer life

Will keep me from dying

Cost (Cost/Insurance, inconvenience/sedations, location, time/duration of test)

Is not expensive

Does not require me to be sedated

Is something I do not have to cancel work to complete

Is cheap

Is quick

Is affordable

Is expensive

Is costly

Is convenient

Is covered by my insurance provider

Can be done at a nearby doctor’s office

Can be done at my doctor’s office

Recommendations

Is recommended by my physician

Is suggested by my health plan

Is endorsed by my hospital

Is encouraged by my family

Is approved by my significant other

Is suggested by my child

Is encouraged by someone on TV

Is endorsed by an advertisement

Is a test my family member or friend has had before

Is socially acceptable

Is for a disease that I hear about in the news

Is recommended by an unbiased panel of expert physicians

Familiarity/Simplicity/Frequency

Is familiar

Is for a disease I have heard of

Is a test I’ve heard of before

Is a test I am well informed about

Is a test I understand well

Is a common, routine procedure

Does not have to be done frequently

Is widely available

Is trustworthy

Has been around for a while

Personal Preferences/Fit

Fits with my religious or spiritual beliefs

Is a natural test

Is a new test

Fits with my values

Is consistent with my personal preferences

Action

Is one I will be scared if I don’t take

Makes me feel that I am taking action toward better health

Is better than no screening at all

Is one I am afraid I will regret if I don’t get

Will make me feel prepared

Will make me feel I have done something proactive

Disease Attributes

Is for a disease with serious consequences

Is for a preventable disease

Is for a disease I am likely to have

Is for a curable disease

Is for a treatable disease

Is for a disease I am showing symptoms of

Is for a widespread disease

**Appendix IV**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Descriptive Statistics for Disease EFA | | | | |
| Item | Range | Mean | SD | Skew |
| Is a test my family member of friend has had before | 0-5 | 2.21 | 1.61 | 0.05 |
| Is a test I am well informed about | 0-6 | 4.73 | 1.30 | -1.01 |
| Is recommended by my physician | 0-6 | 4.80 | 1.24 | -1.06 |
| Is for a treatable disease | 0-6 | 4.52 | 1.46 | -0.94 |
| Will not cause serious problems | 0-6 | 4.88 | 1.22 | -1.16 |
| Is a natural test | 0-6 | 3.40 | 1.73 | -0.24 |
| Fits with my values | 0-6 | 3.66 | 1.74 | -0.44 |
| Will tell me that I have the disease if I do, in fact, have the disease | 0-6 | 5.17 | 1.15 | -1.39 |
| Is trustworthy | 0-6 | 5.21 | 1.06 | -1.42 |
| Will reassure me that I am healthy | 0-6 | 4.48 | 1.53 | -0.94 |
| Will reduce my chances of dying from the disease | 1-6 | 5.06 | 1.24 | -1.38 |
| Will ensure that I won't be surprised by a diagnosis later in life | 1-6 | 4.96 | 1.18 | -1.15 |
| Is costly | 0-6 | 2.39 | 1.65 | 0.33 |
| Is endorsed by an advertisement | 0-6 | 1.94 | 1.79 | 0.63 |
| Will not tell me I do not have the disease if I do have the disease | 0-6 | 3.88 | 1.68 | -0.63 |
| Will not make me worry that I am unhealthy | 0-6 | 3.55 | 1.74 | -0.40 |
| Does not seem scary | 0-6 | 3.10 | 1.83 | -0.11 |
| Is socially acceptable | 0-6 | 4.56 | 1.44 | -0.95 |
| Is familiar | 0-6 | 3.58 | 1.57 | -0.42 |
| Will not falsely diagnose me as having the disease if I do not have the disease | 0-6 | 5.16 | 1.17 | -1.51 |
| Does not have side-effects | 0-6 | 4.30 | 1.40 | -0.73 |
| Is cheap | 0-6 | 3.81 | 1.63 | -0.53 |
| Is for a disease I am showing symptoms of | 0-6 | 4.66 | 1.35 | -1.21 |
| Will protect me from life-threatening disease | 0-6 | 5.06 | 1.22 | -1.32 |
| Will catch the disease early | 0-6 | 5.14 | 1.17 | -1.57 |
| Correctly classifies people who truly do NOT have the disease | 0-6 | 5.11 | 1.18 | -1.21 |
| Will save my life | 1-6 | 5.16 | 1.19 | -1.36 |
| Will give me consistent results (i.e. if I have the disease and am tested for it twice, both tests will tell me that I have the disease) | 1-6 | 5.24 | 1.08 | -1.41 |
| Has no risk of pain | 0-6 | 3.76 | 1.55 | -0.38 |
| Fits with my religious or spiritual beliefs | 0-6 | 2.99 | 2.00 | -0.11 |
| Is quick | 0-6 | 3.39 | 1.74 | -0.22 |
| Is affordable | 0-6 | 4.02 | 1.57 | -0.51 |
| Will result in minimal discomfort | 0-6 | 3.88 | 1.59 | -0.40 |
| Is a common, routine procedure | 0-6 | 3.96 | 1.57 | -0.55 |
| Is safe | 0-6 | 5.07 | 1.16 | -1.50 |
| Is for a preventable disease | 0-6 | 4.74 | 1.37 | -1.11 |
| Is for a curable disease | 0-6 | 4.75 | 1.38 | -1.15 |
| Will make me feel I have done something proactive | 0-6 | 4.25 | 1.41 | -0.73 |
| Will accurately determine the severity of the disease if it is detected | 0-6 | 5.03 | 1.15 | -1.16 |
| Is consistent with my personal preferences | 0-6 | 3.87 | 1.57 | -0.54 |
| Is suggested by my health plan | 0-6 | 4.23 | 1.44 | -0.77 |
| Is approved by my significant other | 0-6 | 3.51 | 1.73 | -0.27 |
| Is for a disease with serious consequences | 0-6 | 4.40 | 1.49 | -0.85 |
| Is recommended by an unbiased panel of expert physicians | 0-6 | 4.68 | 1.38 | -1.11 |
| Is for a disease I am likely to have | 0-6 | 4.77 | 1.31 | -1.09 |
| Is a test I’ve heard of before | 0-6 | 3.29 | 1.63 | -0.24 |
| Is better than no screening at all | 0-6 | 4.49 | 1.38 | -0.78 |
| Will increase my life expectancy | 0-6 | 4.85 | 1.34 | -1.16 |
| Will not tell me that I have the disease when I actually do not have the disease | 0-6 | 4.49 | 1.87 | -1.13 |
| Will not lead to psychological distress | 0-6 | 4.42 | 1.39 | -0.77 |
| Has been shown to lead to improved health outcomes | 0-6 | 4.91 | 1.21 | -1.23 |
| Is a simple procedure | 0-6 | 3.97 | 1.45 | -0.46 |
| Is encouraged by my family | 0-6 | 4.14 | 1.48 | -0.67 |
| Is expensive | 0-6 | 2.27 | 1.62 | 0.44 |
| Will result in only minor side effects | 0-6 | 3.92 | 1.47 | -0.45 |
| Is endorsed by my hospital | 0-6 | 3.95 | 1.59 | -0.60 |
| Is a test I understand well | 0-6 | 4.45 | 1.48 | -0.92 |
| Is one I am afraid I will regret if I don’t get | 0-6 | 4.05 | 1.44 | -0.66 |
| Will be a comfortable experience | 0-6 | 4.15 | 1.48 | -0.64 |
| Distinguishes between diseased and non-diseased people | 0-6 | 4.84 | 1.31 | -1.01 |
| Will not lead to complications with the procedure | 0-6 | 4.81 | 1.19 | -1.01 |
| Is one that can be done at a nearby doctor’s office | 0-6 | 3.86 | 1.59 | -0.50 |
| Has no risk of anxiety | 0-6 | 3.84 | 1.54 | -0.42 |
| Does not have to be done frequently | 0-6 | 4.15 | 1.47 | -0.64 |
| Is suggested by my child | 0-6 | 2.11 | 1.81 | 0.47 |
| Is for a disease I have heard of | 0-6 | 3.35 | 1.77 | -0.33 |
| Makes me feel that I am taking action toward better health | 0-6 | 4.47 | 1.35 | -0.85 |
| Is one I will be scared if I don’t take | 0-6 | 3.88 | 1.48 | -0.45 |
| Gives me accurate information about whether or not I have the disease | 1-6 | 5.16 | 1.17 | -1.38 |
| Makes me feel safer | 0-6 | 4.86 | 1.26 | -1.17 |
| Is a new test | 0-6 | 2.49 | 1.63 | 0.21 |
| Will find the disease early enough that I can be treated | 1-6 | 4.99 | 1.24 | -1.17 |
| Is covered by my insurance provider | 0-6 | 4.55 | 1.39 | -0.88 |
| Will help me to live a longer life | 0-6 | 5.01 | 1.27 | -1.30 |
| Will tell me that I do not have the disease if I, in fact, do not have the disease | 0-6 | 5.12 | 1.37 | -1.92 |
| Does not require me to be sedated | 0-6 | 3.35 | 1.80 | -0.21 |
| Will improve my health | 0-6 | 4.94 | 1.19 | -1.16 |
| Is something I do not have to cancel work to complete | 0-6 | 3.39 | 1.81 | -0.23 |
| Is one that can be done at my doctor’s office | 0-6 | 3.75 | 1.67 | -0.43 |
| Is convenient | 0-6 | 4.11 | 1.49 | -0.66 |
| Is widely available | 0-6 | 4.18 | 1.50 | -0.70 |
| Is for a widespread disease | 0-6 | 3.87 | 1.62 | -0.57 |
| Is not expensive | 0-6 | 3.81 | 1.70 | -0.36 |
| Will make me feel prepared | 0-6 | 4.55 | 1.28 | -0.87 |
| Is for a disease that I hear about in the news | 0-6 | 2.91 | 1.89 | 0.03 |
| Will keep me from dying | 0-6 | 5.12 | 1.29 | -1.46 |
| Is encouraged by someone on TV | 0-6 | 1.84 | 1.71 | 0.66 |
| Is accurate at identifying people who truly have the disease | 1-6 | 5.15 | 1.15 | -1.30 |
| Will not make me worried unnecessarily | 0-6 | 4.34 | 1.53 | -0.89 |
| Has been around for a while | 0-6 | 3.57 | 1.63 | -0.41 |
| Will not lead to unnecessary treatment | 0-6 | 4.70 | 1.40 | -1.13 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Final model statistics for reduced scale for the disease items | | | | | | |
| Factor Model | CFI | TLI | RMSEA | RMSR | Iterations | Total Items |
| 5-Factor Reduced | 0.979 | 0.961 | 0.046 | 0.02 | 5 | 21 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Final scale for the disease items | | | | | |
| Factor | Item | Standardized Loading | h2 | u2 | com |
| Accuracy | Will tell me that I have the disease if I do, in fact, have the disease (8) | 0.78 | 0.62 | 0.38 | 1 |
| Is trustworthy (9) | 0.64 | 0.56 | 0.44 | 1.1 |
| Will not falsely diagnose me as having the disease if I do not have the disease (2) | 0.84 | 0.62 | 0.38 | 1 |
| Correctly classifies people who truly do NOT have the disease (8) | 0.79 | 0.6 | 0.4 | 1 |
| Will give me consistent results (i.e. if I have the disease and am tested for it twice, both tests will tell me that I have the disease) (10) | 0.81 | 0.7 | 0.3 | 1 |
| Gives me accurate information about whether or not I have the disease (13) | 0.72 | 0.58 | 0.42 | 1 |
| Is accurate at identifying people who truly have the disease (13) | 0.67 | 0.59 | 0.41 | 1.1 |
| Benefit | Will save my life (9) | 0.69 | 0.63 | 0.37 | 1.1 |
| Will increase my life expectancy (11) | 0.82 | 0.68 | 0.32 | 1 |
| Will help me to live a longer life (18) | 0.73 | 0.58 | 0.42 | 1 |
| Will keep me from dying (11) | 0.69 | 0.56 | 0.44 | 1.1 |
| Familiarity | Is a test my family member of friend has had before (1) | 0.73 | 0.51 | 0.49 | 1 |
| Is a test I’ve heard of before (9) | 0.82 | 0.64 | 0.36 | 1 |
| Is for a disease I have heard of (10) | 0.68 | 0.55 | 0.45 | 1.1 |
| Is for a disease that I hear about in the news (10) | 0.7 | 0.56 | 0.44 | 1.1 |
| Monetary Cost | Is cheap (4) | 0.9 | 0.76 | 0.24 | 1 |
| Is affordable (14) | 0.79 | 0.71 | 0.29 | 1 |
| Is not expensive (8) | 0.71 | 0.59 | 0.41 | 1 |
| Personal Preference | Fits with my values (7) | 0.68 | 0.51 | 0.49 | 1.1 |
| Will not make me worry that I am unhealthy (16) | 0.34 | 0.48 | 0.52 | 3.3 |
| Is consistent with my personal preferences (3) | 0.82 | 0.73 | 0.27 | 1 |

**Appendix V**

**Summary:**

Oblique rotation methods such as Oblimin and Quartimin rotation support a 5-Factor solution. While final items and loading vary slightly from model to model, the same basic 5-Factors relating to how familiar the individual is with the screening test, the accuracy of the test, the benefit of the test, cost, and personal fit/preferences repeatedly arise. The GeominQ rotation returns a 4-Factor structure that retained all factors from the 5-Factor model except for personal fit/preference.

Orthogonal methods returned fewer factors. The Promax rotation returned 3-Factors while the Varimax and Quartimax rotations returns 2-Factor solutions. Given the factors that are extracted, I suspect that the inability of most of these orthogonal methods to successfully extract a 5-Factor solution like the oblique methods did is due to the fact that many of these factors are correlated with one another. For instance, the benefit and accuracy factors should reasonably be correlated as in real life the accuracy of a test is related to how much benefit can be produced by the test.

When testing bi-factor solutions the Bifactor-Quartimin rotation failed to produce any solutions, while the Schmid-Leiman transformation returned a general factor and 4 additional factors. This solution generally supported a new factor regarding the comfort of the test, an accuracy factor and a benefit factor similar to our previous factors, as well as a factor that looked like a mix of our previously discussed familiarity and fit/preference factors.

Details regarding each rotation method, their models and fits, can be seen below. All rotation steps and final items can be found at [osf.io/agx3p](file:///C:\Users\shafferv\Library\Group%20Containers\3L68KQB4HG.group.com.readdle.smartemail\databases\messagesData\1\25730\osf.io\agx3p). Taken together, we will support the idea of the 5-Factor solution that was identified by the initial Oblimin rotation.

**Oblique Methods:**

When analyzing the data with alternative rotation methods, variations in factors and items were found. Using the quartimin rotation, we find that 8, 7, 6, and 3 factor models do not fit as almost all of these models end up with too few items on the last factor, and for the 3-factor model the model splinters at the second iteration. The 5-Factor model fit well, and consisted of the same factors (with slightly fewer items) as our reported Oblimin rotation. The 4, 2, and 1 Factor models fit, but the fit of these models was appreciably lower than the fit of the 2-Factor model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model statistics for EFAs under quartimin rotation | | | | | | |
| Factor Model | CFI | TLI | RMSEA | RMSR | Iterations | Total Items |
| 1-Factor | 0.484 | 0.470 | 0.094 | 0.13 | 2 | 85 |
| 2-Factor | 0.749 | 0.734 | 0.068 | 0.05 | 3 | 78 |
| 4-Factor | 0.832 | 0.811 | 0.059 | 0.04 | 3 | 76 |
| 5-Factor | 0.865 | 0.840 | 0.057 | 0.04 | 4 | 67 |

Using the geominQ rotation, we find that 8, 7, 6, and 5 factor models do not fit as all of these models end up with too few items on the last factor. The 4-Factor model fit well, and retained all of our original factors except for fit. The 3, 2, and 1 Factor models fit, but the fit of these models was appreciably lower than the fit of the 4-Factor model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model statistics for EFAs under geominQ rotation | | | | | | |
| Factor Model | CFI | TLI | RMSEA | RMSR | Iterations | Total Items |
| 1-Factor | 0.484 | 0.470 | 0.094 | 0.13 | 2 | 85 |
| 2-Factor | 0.754 | 0.739 | 0.068 | 0.05 | 3 | 76 |
| 3-Factor | 0.789 | 0.769 | 0.066 | 0.05 | 3 | 73 |
| 4-Factor | 0.839 | 0.816 | 0.061 | 0.04 | 5 | 66 |

**Orthogonal Methods:**

Using the promax rotation, we find that 8, 7, 6, 5, and 4 factor models do not fit as all of these models end up with too few items on the last factor. The 3-Factor model fit well, and consisted of one factor regarding benefit and accuracy, a second regarding comfort, and a final factor about familiarity. The 2 and 1 Factor models fit, but the fit of these models was appreciably lower than the fit of the 3-Factor model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model statistics for EFAs under promax rotation | | | | | | |
| Factor Model | CFI | TLI | RMSEA | RMSR | Iterations | Total Items |
| 1-Factor | 0.484 | 0.470 | 0.094 | 0.13 | 2 | 85 |
| 2-Factor | 0.714 | 0.739 | 0.069 | 0.05 | 2 | 84 |
| 3-Factor | 0.802 | 0.781 | 0.065 | 0.05 | 7 | 69 |

Using the varimax rotation, we find that 8, 7, 6, 5, 4, and 3 factor models do not fit as almost all of these models end up with too few items on the last factor, and for the 3-factor model the model splinters at the second iteration. The 2-Factor model fit well, and consisted of one factor regarding benefit and accuracy, a second regarding comfort and convenience. The 1 Factor model fit, but the fit of these models was appreciably lower than the fit of the 2-Factor model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model statistics for EFAs under varimax rotation | | | | | | |
| Factor Model | CFI | TLI | RMSEA | RMSR | Iterations | Total Items |
| 1-Factor | 0.484 | 0.470 | 0.094 | 0.13 | 2 | 85 |
| 2-Factor | 0.757 | 0.742 | 0.07 | 0.05 | 2 | 71 |

Using the quartimax rotation, we find that 8, 7, 6, 5, 4, and 3 factor models do not fit as all of these models end up with too few items on the last factor. The 2-Factor model fit well, and consisted of one factor regarding benefit and accuracy, a second regarding comfort and convenience. The 1 Factor model fit, but the fit of these models was appreciably lower than the fit of the 2-Factor model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model statistics for EFAs under quartimax rotation | | | | | | |
| Factor Model | CFI | TLI | RMSEA | RMSR | Iterations | Total Items |
| 1-Factor | 0.484 | 0.470 | 0.094 | 0.13 | 2 | 85 |
| 2-Factor | 0.756 | 0.741 | 0.069 | 0.05 | 2 | 74 |

**Bifactor Models**

Using the Schmid-Leiman transformation, and an oblique Oblimin rotation, only a solution consisting of a general factor and 4 additional factors fit. However, when attempting to interpret the factors it was difficult to identify meaningful factor categories given the items in each factor. Unlike the various rotations methods used, this transformation does not result in a multitude of fit statistics being readily apparent, but the RMSR is 0.04.

Using the Bifactor-Quartimin oblique rotation, no models fit. All models either failed to converge (8-5 factor models), or broke down at a second or third iteration (4-2 factor models).

**Appendix VI**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Descriptive Statistics for EFA Study | | | | |
| Item | Range | Mean | SD | Skew |
| Is a test my family member of friend has had before | 0-6 | 3.099 | 1.822 | -0.141 |
| Is a test I am well informed about | 0-6 | 4.720 | 1.395 | -1.187 |
| Is recommended by my physician | 0-6 | 4.795 | 1.293 | -1.193 |
| Is for a treatable disease | 0-6 | 4.605 | 1.484 | -1.162 |
| Will not cause serious problems | 0-6 | 4.880 | 1.291 | -1.181 |
| Is a natural test | 0-6 | 3.240 | 1.761 | -0.123 |
| Fits with my values | 0-6 | 3.592 | 1.740 | -0.383 |
| Will tell me that I have the disease if I do, in fact, have the disease | 0-6 | 5.139 | 1.321 | -1.795 |
| Is trustworthy | 0-6 | 5.197 | 1.113 | -1.541 |
| Will reassure me that I am healthy | 0-6 | 4.517 | 1.568 | -1.038 |
| Will reduce my chances of dying from the disease | 0-6 | 5.045 | 1.233 | -1.350 |
| Will ensure that I won't be surprised by a diagnosis later in life | 0-6 | 4.955 | 1.215 | -1.207 |
| Is costly | 0-6 | 2.443 | 1.725 | 0.334 |
| Is endorsed by an advertisement | 0-6 | 1.864 | 1.838 | 0.759 |
| Will not tell me I do not have the disease if I do have the disease | 0-6 | 3.976 | 1.709 | -0.744 |
| Will not make me worry that I am unhealthy | 0-6 | 3.560 | 1.771 | -0.321 |
| Does not seem scary | 0-6 | 3.035 | 1.921 | -0.045 |
| Is socially acceptable | 0-6 | 4.619 | 1.433 | -1.140 |
| Is familiar | 0-6 | 3.712 | 1.706 | -0.447 |
| Will not falsely diagnose me as having the disease if I do not have the disease | 0-6 | 5.229 | 1.157 | -1.601 |
| Does not have side-effects | 0-6 | 4.227 | 1.425 | -0.639 |
| Is cheap | 0-6 | 3.597 | 1.710 | -0.287 |
| Is for a disease I am showing symptoms of | 0-6 | 4.715 | 1.313 | -1.298 |
| Will protect me from life-threatening disease | 1-6 | 5.123 | 1.097 | -1.163 |
| Will catch the disease early | 0-6 | 5.283 | 1.032 | -1.618 |
| Correctly classifies people who truly do NOT have the disease | 1-6 | 5.120 | 1.188 | -1.339 |
| Will save my life | 0-6 | 5.229 | 1.206 | -1.666 |
| Will give me consistent results (i.e. if I have the disease and am tested for it twice, both tests will tell me that I have the disease) | 1-6 | 5.277 | 1.012 | -1.452 |
| Has no risk of pain | 0-6 | 3.576 | 1.690 | -0.287 |
| Fits with my religious or spiritual beliefs | 0-6 | 2.939 | 2.024 | -0.009 |
| Is quick | 0-6 | 3.163 | 1.789 | -0.127 |
| Is affordable | 0-6 | 4.077 | 1.578 | -0.563 |
| Will result in minimal discomfort | 0-6 | 3.859 | 1.607 | -0.387 |
| Is a common, routine procedure | 0-6 | 3.933 | 1.630 | -0.580 |
| Is safe | 0-6 | 5.096 | 1.138 | -1.446 |
| Is for a preventable disease | 0-6 | 4.717 | 1.370 | -1.064 |
| Is for a curable disease | 0-6 | 4.821 | 1.301 | -1.092 |
| Will make me feel I have done something proactive | 0-6 | 4.392 | 1.358 | -0.949 |
| Will accurately determine the severity of the disease if it is detected | 0-6 | 5.163 | 1.071 | -1.393 |
| Is consistent with my personal preferences | 0-6 | 3.963 | 1.612 | -0.540 |
| Is suggested by my health plan | 0-6 | 4.181 | 1.551 | -0.792 |
| Is approved by my significant other | 0-6 | 3.549 | 1.728 | -0.375 |
| Is for a disease with serious consequences | 0-6 | 4.552 | 1.390 | -0.962 |
| Is recommended by an unbiased panel of expert physicians | 0-6 | 4.701 | 1.380 | -1.056 |
| Is for a disease I am likely to have | 0-6 | 4.848 | 1.335 | -1.357 |
| Is a test I’ve heard of before | 0-6 | 3.381 | 1.768 | -0.259 |
| Is better than no screening at all | 0-6 | 4.659 | 1.349 | -1.153 |
| Will increase my life expectancy | 0-6 | 5.016 | 1.243 | -1.263 |
| Will not tell me that I have the disease when I actually do not have the disease | 0-6 | 4.600 | 1.844 | -1.277 |
| Will not lead to psychological distress | 0-6 | 4.400 | 1.453 | -0.881 |
| Has been shown to lead to improved health outcomes | 0-6 | 4.984 | 1.154 | -1.375 |
| Is a simple procedure | 0-6 | 3.907 | 1.604 | -0.535 |
| Is encouraged by my family | 0-6 | 4.133 | 1.610 | -0.719 |
| Is expensive | 0-6 | 2.341 | 1.693 | 0.418 |
| Will result in only minor side effects | 0-6 | 3.987 | 1.519 | -0.635 |
| Is endorsed by my hospital | 0-6 | 3.979 | 1.635 | -0.735 |
| Is a test I understand well | 0-6 | 4.531 | 1.416 | -1.044 |
| Is one I am afraid I will regret if I don’t get | 0-6 | 4.253 | 1.536 | -0.918 |
| Will be a comfortable experience | 0-6 | 4.032 | 1.613 | -0.673 |
| Distinguishes between diseased and non-diseased people | 0-6 | 4.992 | 1.325 | -1.422 |
| Will not lead to complications with the procedure | 0-6 | 4.824 | 1.261 | -1.177 |
| Is one that can be done at a nearby doctor’s office | 0-6 | 3.672 | 1.714 | -0.372 |
| Has no risk of anxiety | 0-6 | 3.728 | 1.690 | -0.373 |
| Does not have to be done frequently | 0-6 | 4.032 | 1.532 | -0.570 |
| Is suggested by my child | 0-6 | 2.251 | 1.862 | 0.296 |
| Is for a disease I have heard of | 0-6 | 3.429 | 1.788 | -0.369 |
| Makes me feel that I am taking action toward better health | 0-6 | 4.672 | 1.269 | -0.934 |
| Is one I will be scared if I don’t take | 0-6 | 4.048 | 1.611 | -0.670 |
| Gives me accurate information about whether or not I have the disease | 0-6 | 5.203 | 1.168 | -1.500 |
| Makes me feel safer | 0-6 | 4.885 | 1.281 | -1.216 |
| Is a new test | 0-6 | 2.517 | 1.610 | 0.153 |
| Will find the disease early enough that I can be treated | 0-6 | 5.128 | 1.137 | -1.349 |
| Is covered by my insurance provider | 0-6 | 4.525 | 1.453 | -1.041 |
| Will help me to live a longer life | 0-6 | 5.168 | 1.097 | -1.351 |
| Will tell me that I do not have the disease if I, in fact, do not have the disease | 0-6 | 5.168 | 1.359 | -2.014 |
| Does not require me to be sedated | 0-6 | 3.109 | 1.850 | -0.064 |
| Will improve my health | 0-6 | 5.013 | 1.194 | -1.352 |
| Is something I do not have to cancel work to complete | 0-6 | 3.152 | 1.862 | -0.167 |
| Is one that can be done at my doctor’s office | 0-6 | 3.669 | 1.737 | -0.469 |
| Is convenient | 0-6 | 3.896 | 1.672 | -0.641 |
| Is widely available | 0-6 | 4.192 | 1.513 | -0.761 |
| Is for a widespread disease | 0-6 | 3.904 | 1.647 | -0.681 |
| Is not expensive | 0-6 | 3.811 | 1.708 | -0.485 |
| Will make me feel prepared | 0-6 | 4.536 | 1.328 | -0.956 |
| Is for a disease that I hear about in the news | 0-6 | 2.792 | 1.864 | 0.061 |
| Will keep me from dying | 0-6 | 5.109 | 1.263 | -1.619 |
| Is encouraged by someone on TV | 0-6 | 1.888 | 1.838 | 0.582 |
| Is accurate at identifying people who truly have the disease | 1-6 | 5.131 | 1.152 | -1.343 |
| Will not make me worried unnecessarily | 0-6 | 4.304 | 1.533 | -0.832 |
| Has been around for a while | 0-6 | 3.507 | 1.645 | -0.373 |
| Will not lead to unnecessary treatment | 0-6 | 4.808 | 1.306 | -1.122 |

**Appendix VII**

../Screening%20Cues%20Questionnaire/EFA-Spring%202017%20Study/December%20Analyses/Scree.all.cancer.data.pdf

Scree plot for undergraduate EFA Study

|  |
| --- |
|  |
| VSS plot for undergraduate EFA Study |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Appendix VIII**  Uncolored cells represent no apriori hypothesis. Red cells represent relationships that did not match hypotheses. Green cells represent relationships that did match hypotheses. | | | | | | | |
| Measured Variable | Accuracy | Benefit | Cost | Familiarity | Personal Preference | Measurement Scale | M (SD) |
| Age | 0.02 | 0.03 | -0.04 | 0.01 | -0.01 | Number entry | 18.81 (1.447) |
| Gender  Females are more likely to rate accuracy and personal fit as important to them | 0.17\*\*\* | 0.05 | 0 | -0.07 | 0.14\* | 0=male, 1=female | 61.70% female |
| Race  White participants are more likely to rate accuracy and benefit as important to them, whereas non-white participants are more likely to rate cost as important to them. | 0.12\* | 0.12\* | -0.15\*\*\* | -0.1 | 0 | 0=nonwhite 1=white | 81.23% white |
| Subjective numeracy  Individuals with higher numeracy are more concerned with the accuracy of a test. | 0.22\* | 0.18 | 0.09 | -0.07 | -0.01 | 1-6 scale | 4.220 (0.930) |
| Maximizer-Minimizer  Medical maximizers are more concerned with the benefit received from the test. | 0.14 | 0.25\*\*\* | -0.03 | 0.09 | 0.12 | 1-7 scale | 4.394 (0.881) |
| BMQ harm | 0.16 | 0.1 | 0 | -0.17 | -0.12 | 1-5 scale | 3.612 (0.375) |
| BMQ overuse | 0.01 | 0.05 | -0.06 | -0.06 | -0.05 | 1-5 scale | 3.271 (0.451) |
| DOSPERT social | 0.05 | -0.06 | -0.08 | -0.17 | -0.14 | 1-7 scale | 4.682 (0.791) |
| DOSPERT health  Individuals who are more risk taking in health are less likely to be concerned with the accuracy of the test. | -0.24\*\*\* | -0.13 | -0.08 | 0.02 | -0.12 | 1-7 scale | 3.514 (1.216) |
| DOSPERT recreation | -0.08 | -0.05 | -0.01 | -0.07 | -0.01 | 1-7 scale | 3.985 (1.469) |
| DOSPERT financial  Individuals who are more risk taking in finance are less likely to be concerned with the accuracy of the test. | -0.31\*\*\* | -0.19 | -0.17 | 0.06 | -0.05 | 1-7 scale | 2.974 (1.320) |
| DOSPERT ethical  Individuals who are more risk taking in ethical situations are less likely to be concerned with the accuracy of the test or the benefit of the test. | -0.39\*\*\* | -0.26\*\*\* | -0.03 | 0.13 | -0.11 | 1-7 scale | 2.598 (1.123) |
| Cancer anxiety scale | 0.04 | 0.08 | 0.03 | 0.08 | 0.1 | 1-5 scale | 3.258 (1.033) |
| MHLC internal | 0.2 | 0.21 | 0.05 | -0.05 | 0.02 | 1-5 scale | 3.217 (0.891) |
| MHLC powerful others | -0.04 | 0.02 | 0 | 0 | 0.07 | 1-5 scale | 2.675 (0.979) |
| MHLC chance | -0.07 | 0.01 | 0.02 | 0.03 | -0.06 | 1-5 scale | 2.745 (0.882) |
| AA-med  Those who are more averse to ambiguity are more likely to be concerned with personal fit. | 0.16 | 0.15 | 0.14 | 0.05 | 0.25\*\*\* | 1-5 scale | 3.234 (0.656) |
| MIHT cognitive | -0.05 | -0.01 | 0.16 | 0.12 | -0.01 | 1-5 scale | 3.484 (1.646) |
| MIHT behavioral  Those higher in the need for reassurance of their health status are more concerned with the benefit of the test. | 0.15 | 0.22\* | 0.05 | 0.1 | 0.15 | 1-5 scale | 3.392 (1.021) |
| MIHT perceptual  Those who are more aware of their health are more concerned with the accuracy and benefit of the screening test. | 0.41\*\*\* | 0.27\*\*\* | 0.05 | -0.07 | 0.1 | 1-5 scale | 3.904 (0.638) |
| MIHT affect | 0.07 | 0.17 | 0.02 | 0.18 | 0.13 | 1-5 scale | 3.017 (0.873) |
| FACIT meaning  Higher scores in spiritual meaning were related to rating accuracy and benefit as more important. | 0.24\*\*\* | 0.27\*\*\* | -0.06 | -0.03 | 0.18 | 0-16 score | 12.04 (3.756) |
| FACIT peace  Higher scores in feelings of peace were related to rating benefit as more important. | 0.06 | 0.25\*\*\* | 0.02 | 0.1 | 0.14 | 0-16 score | 10.05 (3.100) |
| FACIT faith  Higher scores in faith were related to rating personal fit as more important. | -0.02 | 0.03 | 0.05 | 0.12 | 0.48\*\*\* | 0-8 score | 4.393 (3.432) |
| FACIT  Those with higher scores in spiritual well being were more concerned with the benefit of the test as well as how it fits with their personal preferences | 0.13 | 0.24\*\*\* | 0 | 0.08 | 0.32\*\*\* | 0-40 score | 26.46 (8.874) |
| AURO internal  Those with more internal religious orientations were more concerned with personal fit. | -0.04 | 0.01 | 0.02 | 0.07 | 0.43\*\*\* | 1-7 scale | 3.949 (1.682) |
| AURO external  Those with more external religious orientations were more concerned with personal fit. | -0.13 | -0.03 | 0.01 | 0.12 | 0.4\*\*\* | 1-7 scale | 3.560 (1.596) |
| Religiosity  Those scoring higher on religiosity were more concerned with personal fit. | -0.02 | -0.05 | 0.05 | 0.04 | 0.42\*\*\* | 1-3 scale | 2.185 (1.180) |
| Value of Health | -0.2 | -0.06 | 0.01 | 0.04 | -0.02 | 1-4 scale | 3.970 (0.726) |
| HOS information | -0.03 | -0.05 | 0.13 | 0.08 | 0.07 | 0-7 | 3.567 (0.527) |
| HOS behavior | -0.03 | -0.05 | 0.13 | 0.08 | 0.07 | 0-9 | 4.701 (1.581) |
| Big 5 openness | 0.14 | 0.1 | -0.02 | -0.07 | -0.11 | 1-5 scale | 3.433 (0.829) |
| Big 5 conscientiousness | 0.1 | 0.07 | -0.12 | -0.14 | 0.09 | 1-5 scale | 3.484 (0.857) |
| Big 5 extraversion | 0.03 | 0.08 | -0.07 | 0.02 | 0.14 | 1-5 scale | 3.341 (0.651) |
| Big 5 agreeableness | 0.14 | 0.15 | 0.02 | 0.03 | 0.05 | 1-5 scale | 3.646 (0.866) |
| Big 5 neuroticism | 0.05 | 0.01 | 0.01 | 0.03 | -0.02 | 1-5 scale | 3.075 (1.024) |
| IE control | 0.03 | 0.03 | 0.05 | 0.07 | -0.02 | Sum between 0-23 | 12.1 (2.804) |
| Do you avoid or put off going to the doctor because you cannot afford it?  Those who more frequently put off visiting the doctor due to cost are also more concerned with the cost of a screening test | -0.04 | -0.2 | 0.29\*\*\* | -0.02 | 0.01 | 1 (no, never)-7 (yes, all the time) | 2.026 (0.707) |
| How much does cancer scare you?  Those who are more scared of cancer are more concerned with the accuracy and benefit of the test. | 0.24\*\*\* | 0.28\*\*\* | 0.06 | 0.07 | 0.16 | 1 (not at all)-7(extremely) | 5.437 (1.453) |
| When you think about the possibility of having cancer, how worried does it make you feel?  Those who are more worried by cancer are more concerned with the benefit of the test. | 0.2 | 0.26\*\*\* | 0.12 | 0.13 | 0.16 | 1 (not at all)-7(extremely) | 5.026 (1.692) |
| How likely do you think it is that you will get cancer one day? | 0.18 | 0.1 | -0.03 | 0.01 | -0.03 | 1 (not at all likely)-7(extremely likely) | 4.105 (1.732) |
| How difficult is it for you to find health insurance that you can afford? | 0.19 | 0.18 | -0.13 | 0.08 | 0.01 | 1 (very difficult)-3 (not at all difficult) | 2.543 (0.333) |
| How difficult is it for you to find a health insurance plan with the type of coverage that you need? | 0.15 | 0.13 | -0.16 | 0.05 | -0.01 | 1 (very difficult)-3 (not at all difficult) | 2.526 (0.333) |
| Would you get more medical care if you could afford it?  Those who state they would get more medical care if they could afford it are also more likely to state that the cost of a screening test is important to them. | -0.05 | -0.01 | 0.25\*\*\* | 0.11 | 0.05 | 1 (no, not at all)-7 (yes, definitely) | 4.626 (2.121) |
| Can you get the medical care that you need?  Those who can get the medical care that they need are more likely to be concerned with the accuracy and benefit of a screening test. | 0.25\*\*\* | 0.32\*\*\* | -0.18 | -0.07 | 0.01 | 1 (no, not at all)-7 (yes, definitely) | 6.307 (1.014) |
| If you get sick or have an accident, how worried are you that you will not be able to pay your medical bills?  Those who are more concerned they may not be able to pay their medical bills should they have an accident are also more concerned with the cost of a screening test. | -0.06 | -0.1 | 0.34\*\*\* | 0.09 | 0.12 | 1 (not at all worried) -3 (very worried) | 1.532 (0.500) |
| How many first degree relatives (i.e. parents or siblings) do you have who have had cancer? | -0.03 | 0 | -0.03 | 0 | -0.01 | 0-20+ | 1.43 (3.333) |
| How many second degree relatives (i.e. aunts, uncles, cousins) do you have who have had cancer? | 0.02 | 0.04 | -0.04 | -0.02 | -0.02 | 0-20+ | 2.992 (2.667) |
| How many friends do you have who have had cancer? | -0.1 | -0.02 | 0.02 | 0.1 | 0.06 | 0-20+ | 2.109 (0.928) |
| Health Access Item/Delay: Once you get there, you have to wait too long to see the doctor | -0.05 | -0.07 | 0.08 | 0.04 | 0 | 0=no, 1=yes | 16.17% yes |
| Health Care Access Item/Afford: Getting a regular health check up  Those who have not had a regular health checkup because they could not afford it are less likely to be concerned with the accuracy and benefit of a screening test. | -0.18\*\*\* | -0.15\*\*\* | 0.04 | 0 | 0 | 0=no, 1=yes | 12.04% yes |
| Have you ever been screened for cancer? | -0.06 | -0.08 | -0.01 | 0.02 | -0.07 | 0=no, 1=yes | 5.14% yes |
| Would you want to be screened for cancer in the future?  Those who want to be screened for cancer in the future are more likely to rate accuracy and cost of the screening test as important. | 0.11\* | 0.08 | 0.12\* | 0.03 | 0.01 | 0=no, 1=yes | 83.8% yes |

**Appendix IX**

**Summary:**

Oblique rotation methods such as Oblimin, Quartimin, and GeominQ all support a 6-Factor solution. While final items and loading vary slightly from model to model, the same basic 6-Factors relating to how familiar the individual is with the screening test, the ease or comfort of receiving the screening, the accuracy of the test, the benefit of the test, a need to take action, and attributes of the disease that is being screened for, repeatedly arise.

When orthogonal methods are utilized, the Promax rotation again identifies the 6-Factor model, while the Varimax rotation returns 4-Factors and the Quartimax returns 3. Similar to the undergraduate sample, I suspect that the inability of most of these orthogonal methods to successfully extract a 6-Factor solution is due to the fact that many of these factors are correlated with one another. For instance, the benefit and accuracy factors should reasonably be correlated as in real life the accuracy of a test is related to how much benefit can be produced by the test.

Further, bifactor solutions using the Schmid-Leiman transformation suggested a general factor and 4 additional factors, while the Bifactor-Quartimin oblique rotation, only the model with a general factor and two sub-factors fit. The four factors from the Schmid-Leiman transformation consisted of ease/comfort of the test, benefit of the test, disease characteristics, and familiarity (though these contained variations of items seen in previous factor structures). The factors suggested by the Bifactor-Quartimin rotation generally reflected the benefits of the test and the accuracy of the test.

Details regarding each rotation method, their models and fits, can be seen below. All rotation steps and final items can be found at [osf.io/agx3p/](file:///C:\Users\shafferv\Library\Group%20Containers\3L68KQB4HG.group.com.readdle.smartemail\databases\messagesData\1\25730\osf.io\agx3p). Taken together, we can support the idea of the 6-Factor solution that was identified by the initial Oblimin rotation.

**Oblique Methods:**

When analyzing the data with alternative rotation methods, variations in factors and items were found. Using the quartimin rotation, we find that 8, and 7 factor models do not fit as all of these models end up with too few items on the last factor. The 6-Factor model fit well, and consisted of the same factors as our reported Oblimin rotation. The 5, 4, 3, 2, and 1 Factor models fit, but the fit of these models was appreciably lower than the fit of the 2-Factor model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model statistics for EFAs under quartimin rotation | | | | | | |
| Factor Model | CFI | TLI | RMSEA | RMSR | Iterations | Total Items |
| 1-Factor | 0.651 | 0.633 | 0.091 | 0.090 | 5 | 44 |
| 2-Factor | 0.705 | 0.686 | 0.068 | 0.070 | 3 | 74 |
| 3-Factor | 0.826 | 0.807 | 0.060 | 0.050 | 3 | 63 |
| 4-Factor | 0.836 | 0.815 | 0.055 | 0.040 | 3 | 76 |
| 5-Factor | 0.860 | 0.836 | 0.053 | 0.040 | 3 | 73 |
| 6-Factor | 0.869 | 0.843 | 0.050 | 0.040 | 4 | 79 |

Using the geominQ rotation, we find that 8 and 7 factor models do not fit as all of these models end up with too few items on the last factor. The 6-Factor model fit well, and consisted of the same factors as our reported Oblimin rotation. The 5, 4, 3, 2, and 1 Factor models fit, but the fit of these models was appreciably lower than the fit of the 2-Factor model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model statistics for EFAs under geominQ rotation | | | | | | |
| Factor Model | CFI | TLI | RMSEA | RMSR | Iterations | Total Items |
| 1-Factor | 0.498 | 0.481 | 0.094 | 0.12 | 4 | 70 |
| 2-Factor | 0.708 | 0.689 | 0.068 | 0.070 | 7 | 73 |
| 3-Factor | 0.820 | 0.794 | 0.065 | 0.050 | 3 | 50 |
| 4-Factor | 0.835 | 0.814 | 0.055 | 0.050 | 3 | 76 |
| 5-Factor | 0.861 | 0.838 | 0.052 | 0.040 | 3 | 76 |
| 6-Factor | 0.874 | 0.847 | 0.051 | 0.040 | 5 | 72 |

**Orthogonal Methods:**

Using the promax rotation, we find that 8 and 7 factor models do not fit as all of these models end up with too few items on the last factor. The 6-Factor model fit well, and consisted of the same factors as our reported Oblimin rotation. The 5, 4, 3, 2, and 1 Factor models fit, but the fit of these models was appreciably lower than the fit of the 2-Factor model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model statistics for EFAs under promax rotation | | | | | | |
| Factor Model | CFI | TLI | RMSEA | RMSR | Iterations | Total Items |
| 1-Factor | 0.498 | 0.481 | 0.091 | 0.120 | 4 | 70 |
| 2-Factor | 0.675 | 0.654 | 0.070 | 0.070 | 2 | 76 |
| 3-Factor | 0.802 | 0.781 | 0.060 | 0.050 | 3 | 68 |
| 4-Factor | 0.842 | 0.821 | 0.054 | 0.040 | 3 | 75 |
| 5-Factor | 0.861 | 0.837 | 0.052 | 0.040 | 3 | 76 |
| 6-Factor | 0.875 | 0.850 | 0.050 | 0.040 | 4 | 76 |

Using the varimax rotation, we find that 8, 7, 6, and 5 factor models do not fit as all of these models end up with too few items on the last factor. The 4-Factor model fit well, and consisted of the 4 factors of familiarity, accuracy, benefit, and ease. The 3, 2, and 1 Factor models fit, but the fit of these models was appreciably lower than the fit of the 4-Factor model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model statistics for EFAs under varimax rotation | | | | | | |
| Factor Model | CFI | TLI | RMSEA | RMSR | Iterations | Total Items |
| 1-Factor | 0.498 | 0.481 | 0.091 | 0.120 | 4 | 70 |
| 2-Factor | 0.709 | 0.690 | 0.067 | 0.060 | 3 | 73 |
| 3-Factor | 0.831 | 0.808 | 0.063 | 0.050 | 4 | 52 |
| 4-Factor | 0.843 | 0.816 | 0.059 | 0.050 | 4 | 57 |

Using the quartimax rotation, we find that 8, 7, 6, 5, and 4 factor models do not fit as all of these models end up with too few items on the last factor. The 3-Factor model fit well, and consisted of the factors of familiarity, benefit, and monetary cost. The 2, and 1 Factor models fit, but the fit of these models was appreciably lower than the fit of the 3-Factor model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model statistics for EFAs under quartimax rotation | | | | | | |
| Factor Model | CFI | TLI | RMSEA | RMSR | Iterations | Total Items |
| 1-Factor | 0.498 | 0.481 | 0.091 | 0.120 | 4 | 70 |
| 2-Factor | 0.711 | 0.692 | 0.067 | 0.070 | 3 | 74 |
| 3-Factor | 0.855 | 0.831 | 0.063 | 0.050 | 3 | 45 |

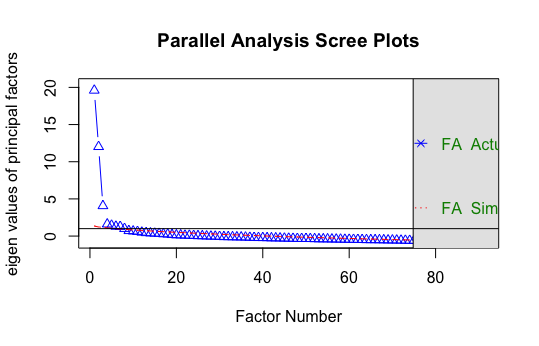
**Bifactor Models**

Using the Schmid-Leiman transformation, and an oblique Oblimin rotation, only a solution consisting of a general factor and 4 additional factors fit. The four factors consisted of ease/comfort of the test, benefit of the test, disease characteristics, and familiarity (though these contained variations of items seen in previous factor structures). Unlike the various rotations methods used, this transformation does not result in a multitude of fit statistics being readily output, but the RMSR is 0.03.

Using the Bifactor-Quartimin oblique rotation, only the model with a general factor and two sub-factors fit. When interpreted, these factors two sub-factors generally reflected the benefits of the test and the accuracy of the test. All other models either failed to converge (8 and 2 factor models), or broke down at a second or third iteration (7-4 factor models).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model statistics for EFAs under varimax rotation | | | | | | |
| Factor Model | CFI | TLI | RMSEA | RMSR | Iterations | Total Items |
| General Factor+2 Sub-Factors | 0.863 | 0.831 | 0.073 | 0.050 | 4 | 33 |

**Appendix X**



Scree plot for Mturk EFA Study

|  |
| --- |
|  |
| VSS plot for Mturk EFA Study |

**Appendix XI**

**Low Quality Data Identification and Handling**

To recruit participants through Amazon Mechanical Turk a Human Intelligence Task (HIT) is created with a set sample size. Mturkers can choose to accept the HIT and complete the survey. Once the sample size was reached, the data were downloaded and then imported into R. This raw data was then investigated for 4 types of low quality data indicators on 5 separate pages of the survey (where a page is one screen of the survey on the webpage that the participant views and responds to).

The four indicators for low quality data consisted of an attention check item, speed of completion of the page, number of clicks counted on the page, and the distribution of responses on the page (Erin M. Buchanan & Scofield, 2018). Code was adapted from Buchanan and Scofield (2018) and can be found at [osf.io/agx3p/](file:///C:\Users\shafferv\Library\Group%20Containers\3L68KQB4HG.group.com.readdle.smartemail\databases\messagesData\1\25730\osf.io\agx3p). Only one page had an attention check item which consisted of the item asking, “Please choose the response labeled 1 on this item” with a 0-6 response scale. Individuals with a response other than 1 on this item were flagged as potentially low quality data.

The speed of completion of the page was tracked with time on page information through the Qualtrics software. A known mean and standard deviation for reading speed was utilized (Trauzettel-Klosinski & Dietz, 2012), and the number of words on each page were tallied. Individuals were flagged as potentially low quality data if they were more than 2 standard deviations below the mean reading time (indicating that they possibly read the items too quickly).

The number of clicks on the page was collected by the Qualtrics software. Ideally, participants would click to indicate their response on each option. Participants were flagged as potentially low quality data if they had less clicks counted on the page than there were items that were responded to on that page. This check is important because automated form fillers, which are easily installed from the internet to complete surveys, do not create clicks on the page when they operate.

The distribution of responses on the page were also investigated. It is worthwhile to note that a statistical test cannot indicate what distribution a set of data is, we can fit the data to multiple distributions and test which distribution provides a better fit. To do this, first a goodness of fit test was calculated where each answer choice was designated as equally likely to simulate a uniform distribution. Next, the goodness of fit was calculated for a normal distribution (see Buchanan & Scofield, 2018). When the uniform distribution fit better when compared to the normal distribution, participants were flagged as potentially low quality data. As automated form fillers are programmed to respond with a uniform distribution, this might indicate an automated form filler was being used.

Once all data were analyzed and coded for these flags, participants were rejected for one of two reasons. First, we wanted to remove participants that we suspected had low quality data. However, as each single indictor provides different, possibly flawed information, it is important to remove participants for being flagged on multiple indictors. For example, an individual may be a fast reader (flagged for speed of completion) and use the keyboard to indicate their responses (flagged for click count), but be paying attention and dutifully completing our survey. Therefore, individuals were additionally removed if they were flagged by two or more of our low quality data indicators on four or more pages. Second, participants were rejected for missing the attention check item. We recognize that this may be a bit redundant (as this is used as one of the flags mentioned above, but as this had been a qualifying piece of information in previous iterations of this study, we felt it worthwhile to include as a straight-forward screening item herein.

Participants were then either approved for the HIT (those who passed the attention check and did not get sufficient flags to be considered low quality data), or rejected (as noted above). Next, the number of individuals who were rejected were then used as the sample size in the second HIT placed on Mturk. This process repeated iteratively until we reached our desired sample size of 300.

**Appendix XII**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Descriptive Statistics for EFA on MTurk Data | | | | |
| Item | Range | Mean | SD | Skew |
| Does not require me to be sedated | 0-6 | 3.12 | 1.95 | -0.16 |
| Will improve my health | 0-6 | 4.70 | 1.40 | -1.04 |
| Is something I do not have to cancel work to complete | 0-6 | 2.72 | 2.07 | 0.11 |
| Is one that can be done at my doctor’s office | 0-6 | 3.69 | 1.71 | -0.54 |
| Is convenient | 0-6 | 3.90 | 1.68 | -0.57 |
| Is widely available | 0-6 | 3.89 | 1.66 | -0.65 |
| Is for a widespread disease | 0-6 | 3.23 | 1.98 | -0.14 |
| Is not expensive | 0-6 | 4.17 | 1.78 | -0.70 |
| Will make me feel prepared | 0-6 | 4.29 | 1.47 | -0.92 |
| Is for a disease that I hear about in the news | 0-6 | 1.92 | 2.01 | 0.67 |
| Will keep me from dying | 0-6 | 5.06 | 1.41 | -1.76 |
| Is encouraged by someone on TV | 0-6 | 1.16 | 1.76 | 1.36 |
| Is accurate at identifying people who truly have the disease | 0-6 | 5.37 | 1.09 | -2.20 |
| Will not make me worried unnecessarily | 0-6 | 4.46 | 1.55 | -1.04 |
| Has been around for a while | 0-6 | 3.02 | 1.78 | -0.18 |
| Will not lead to unnecessary treatment | 0-6 | 4.92 | 1.40 | -1.40 |
| Is trustworthy | 1-6 | 5.34 | 1.04 | -1.65 |
| Is a test my family member of friend has had before | 0-6 | 2.03 | 2.02 | 0.54 |
| Is a test I am well informed about | 0-6 | 4.10 | 1.63 | -0.71 |
| Is recommended by my physician | 0-6 | 4.44 | 1.52 | -1.00 |
| Is for a treatable disease | 0-6 | 3.78 | 1.81 | -0.50 |
| Will not cause serious problems | 0-6 | 4.93 | 1.39 | -1.54 |
| Is a natural test | 0-6 | 2.66 | 1.95 | 0.10 |
| Fits with my values | 0-6 | 3.06 | 2.01 | -0.20 |
| Will tell me that I have the disease if I do, in fact, have the disease | 0-6 | 5.39 | 1.06 | -2.16 |
| Will reassure me that I am healthy | 0-6 | 4.38 | 1.54 | -0.93 |
| Will reduce my chances of dying from the disease | 0-6 | 5.05 | 1.34 | -1.59 |
| Will ensure that I won't be surprised by a diagnosis later in life | 0-6 | 4.52 | 1.62 | -1.17 |
| Is costly | 0-6 | 1.76 | 1.98 | 0.80 |
| Is endorsed by an advertisement | 0-6 | 1.15 | 1.77 | 1.46 |
| Will not make me worry that I am unhealthy | 0-6 | 3.83 | 1.76 | -0.63 |
| Does not seem scary | 0-6 | 3.26 | 1.84 | -0.23 |
| Is socially acceptable | 0-6 | 2.19 | 2.07 | 0.43 |
| Will help other people | 0-6 | 3.53 | 1.98 | -0.50 |
| Is familiar | 0-6 | 2.69 | 1.87 | 0.20 |
| Will not falsely diagnose me as having the disease if I do not have the disease | 0-6 | 5.27 | 1.30 | -2.01 |
| Does not have side-effects | 0-6 | 4.36 | 1.56 | -0.80 |
| Is cheap | 0-6 | 3.57 | 1.89 | -0.32 |
| Is for a disease I am showing symptoms of | 0-6 | 3.71 | 1.88 | -0.53 |
| Will protect me from life-threatening disease | 0-6 | 5.07 | 1.25 | -1.53 |
| Will catch the disease early | 1-6 | 5.23 | 1.10 | -1.70 |
| Correctly classifies people who truly do NOT have the disease | 0-6 | 5.20 | 1.27 | -1.90 |
| Will save my life | 0-6 | 5.16 | 1.21 | -1.75 |
| Will give me consistent results (i.e. if I have the disease and am tested for it twice, both tests will tell me that I have the disease) | 1-6 | 5.34 | 1.01 | -1.60 |
| Has no risk of pain | 0-6 | 3.55 | 1.79 | -0.35 |
| Fits with my religious or spiritual beliefs | 0-6 | 1.64 | 2.06 | 0.87 |
| Is quick | 0-6 | 3.61 | 1.68 | -0.42 |
| Is affordable | 0-6 | 4.65 | 1.51 | -1.12 |
| Will result in minimal discomfort | 0-6 | 4.09 | 1.60 | -0.75 |
| Is a common, routine procedure | 0-6 | 3.72 | 1.69 | -0.45 |
| Is safe | 0-6 | 5.29 | 1.11 | -1.87 |
| Is for a preventable disease | 0-6 | 3.69 | 1.88 | -0.45 |
| Is for a curable disease | 0-6 | 3.68 | 1.86 | -0.45 |
| Will make me feel I have done something proactive | 0-6 | 3.89 | 1.79 | -0.71 |
| Will accurately determine the severity of the disease if it is detected | 0-6 | 5.03 | 1.26 | -1.52 |
| Is consistent with my personal preferences | 0-6 | 3.70 | 1.69 | -0.55 |
| Is suggested by my health plan | 0-6 | 3.51 | 1.88 | -0.48 |
| Is approved by my significant other | 0-6 | 2.69 | 2.06 | 0.05 |
| Is for a disease with serious consequences | 0-6 | 4.30 | 1.57 | -0.84 |
| Is recommended by an unbiased panel of expert physicians | 0-6 | 4.48 | 1.50 | -1.05 |
| Is for a disease I am likely to have | 0-6 | 3.95 | 1.71 | -0.50 |
| Is a test I’ve heard of before | 0-6 | 2.13 | 1.88 | 0.46 |
| Is better than no screening at all | 0-6 | 4.45 | 1.69 | -1.00 |
| Will increase my life expectancy | 0-6 | 4.83 | 1.40 | -1.16 |
| Will not tell me that I have the disease when I actually do not have the disease | 0-6 | 5.03 | 1.59 | -1.82 |
| Will not lead to psychological distress | 0-6 | 4.09 | 1.65 | -0.73 |
| Has been shown to lead to improved health outcomes | 0-6 | 4.91 | 1.30 | -1.40 |
| Is a simple procedure | 0-6 | 3.94 | 1.60 | -0.66 |
| Is encouraged by my family | 0-6 | 2.62 | 2.03 | 0.12 |
| Is expensive | 0-6 | 1.65 | 1.95 | 0.86 |
| Will result in only minor side effects | 0-6 | 4.38 | 1.46 | -0.97 |
| Is endorsed by my hospital | 0-6 | 3.44 | 1.92 | -0.41 |
| Is a test I understand well | 0-6 | 3.73 | 1.86 | -0.53 |
| Is one I am afraid I will regret if I don’t get | 0-6 | 3.27 | 1.96 | -0.25 |
| Will be a comfortable experience | 0-6 | 3.79 | 1.67 | -0.51 |
| Distinguishes between diseased and non-diseased people | 0-6 | 4.95 | 1.39 | -1.56 |
| Will not lead to complications with the procedure | 0-6 | 4.81 | 1.42 | -1.36 |
| Is one that can be done at a nearby doctor’s office | 0-6 | 3.94 | 1.63 | -0.66 |
| Has no risk of anxiety | 0-6 | 3.00 | 1.92 | 0.01 |
| Does not have to be done frequently | 0-6 | 4.14 | 1.59 | -0.79 |
| Is suggested by my child | 0-6 | 1.20 | 1.80 | 1.41 |
| Is for a disease I have heard of | 0-6 | 2.38 | 2.12 | 0.39 |
| Makes me feel that I am taking action toward better health | 0-6 | 4.38 | 1.58 | -1.10 |
| Is one I will be scared if I don’t take | 0-6 | 2.73 | 1.94 | 0.07 |
| Gives me accurate information about whether or not I have the disease | 1-6 | 5.46 | 1.06 | -2.27 |
| Makes me feel safer | 0-6 | 4.42 | 1.55 | -1.08 |
| Is a new test | 0-6 | 1.90 | 1.90 | 0.64 |
| Will find the disease early enough that I can be treated | 0-6 | 5.19 | 1.26 | -1.90 |
| Is covered by my insurance provider | 0-6 | 4.92 | 1.36 | -1.47 |
| Will help me to live a longer life | 0-6 | 5.01 | 1.31 | -1.52 |
| Will tell me that I do not have the disease if I, in fact, do not have the disease | 0-6 | 5.12 | 1.49 | -1.98 |

**Appendix XIII**

**Calculating DCE and TT Thresholds**

Thresholds for DCE were calculated both at the sample level as well as at the individual level for all three risk features. To calculate the DCE threshold at the sample level, first the mean difference between rescaled utilities (provided by the Sawtooth Software) for 1 life saved and 2 lives saved were calculated. Next, the mean difference between the first two levels of each risk feature were calculated. For example, for the false positive risk feature, the rescaled utility for a 25% chance of a false positive was subtracted from the rescaled utility for a 5% chance of a false positive. Next, a ratio was created between the utility gain of 1 life saved and the utility loss of the increased false positive rate. This ratio is then applied to the interval being discussed. For example, for false positives we would take 5 (our base false positive rate), then add the product of 20 (the difference between the first two levels, here 5% and 25%) times the ratio we created. This was done for each risk feature—false positive, false negative, and cost.

The sample level DCE calculations were similar. Each individual’s difference between rescaled utilities for 1 life saved and 2 lives saved were calculated, then the difference between the first two levels of each risk feature were calculated. A ratio was created for each individual that divides utility gained from life saved by utility lost by that risk feature increase. This ratio is then applied to the interval of the risk feature as noted above. This was done for each risk feature—false positive, false negative, and cost.

Sample level TT calculations were based upon responses to branching items. For instance, an individual considering FPs might be asked if they prefer Test A with 1 life saved and 5% risk of FP, or Test B with 2 lives saved and 20% risk of FP. If the person chooses A, they are asked about a lower level of risk, say 2 lives saved and 10% risk, but if they choose B the risk is increased to 40%. Risks are doubled if an individual accepts the higher risk option and split between the last option and the current option considered if they choose the less risky option. This is done iteratively until a person will max out the scale (e.g. reaching 100% chance of FP), or reaches a point of indifference. So, for instance, if an individual accepts a risk of 5%, but rejects a risk of 10%, then rejects a risk of 7%, they would be coded as having a FP threshold of 6. This was done for each risk feature—false positive (max of 100%), false negative (max of 100%), and cost (max of $3200).

**Appendix XIV**

**Undergraduate Low Quality Data Detection**

Three items were utilized to detect low quality data in undergraduate participants. First, data were flagged for completing the survey too quickly. Given the length and complexity of the survey, individuals who took less than 900 seconds (15 minutes) were flagged as low quality data. Second, when introducing participants to the DCE a number of practice items occur. The first of these practice items is what is commonly referred to as a “bozo” question, where two options are presented, one feature is listed under each option, and one option is clearly superior to the other. For our case, this was an option where we asked people if they wanted Test A, where 1 life per 1,000 people screened (0.1%) were saved, or Test B, where 5 lives per 1,000 people screened (0.5%) were saved. As all participants should choose Test B, anyone choosing Test A were flagged as potentially low quality data. Finally, individuals were asked an attention check item which consisted of a question asking, “Please choose the response labeled 1 on this item” with a 0-6 response scale. Individuals with a response other than 1 on this item were flagged as potentially low quality data.

VITA

KD Valentine was born in Merriam, Kansas, where she was raised by her parents alongside her older brother and sister. Growing up she had more interactions with the medical community than most people as she required many surgeries for issues related to her eye (strabismus and astigmatism), and in total has had 5 surgeries for these problems. She received her Bachelor’s degree at Missouri State University where she found her love for statistics and research design. Following her graduation, she immediately began her Master’s degree in Experimental Psychology at the behest of Dr. Harry Hom and Dr. Erin Buchanan. After achieving this degree, she taught introductory statistics for a year and then moved to Columbia, MO to pursue her PhD in Quantitative Psychology. Here she adopted her two dogs, Archie and Warlock.

1. Data was additionally analyzed with other oblique rotations (quartimin and geominQ), some orthogonal rotations (promax, verimax, and quartimax), and bifactor translation /rotation methods (Schmid-Leiman and Bifactor Quartimin, respectively). Results of these analyses and justification of the use of the oblimin rotation presented here are included in Appendix V. [↑](#footnote-ref-1)
2. Data was additionally analyzed with other oblique rotations (quartimin and geominQ), some orthogonal rotations (promax, verimax, and quartimax), and bifactor translation /rotation methods (Schmid-Leiman and Bifactor Quartimin, respectively). Results of these analyses are included in Appendix IX. [↑](#footnote-ref-2)
3. These attributes were chosen to mirror the subscales of ESCAPE as closely as possible. As it was not possible to make varying levels of the “personal preference” subscale for this design, it was left out. [↑](#footnote-ref-3)
4. As before, we attempted to match these attributes as closely as possible to those of ESCAPE. As the familiarity and personal preference subscales of ESCAPE could not be made into the necessary numeric scales for the TT, they were left out. [↑](#footnote-ref-4)
5. It is worthwhile to mention that these results are the same regardless of whether all values predict choice in a single simultaneous equation, each set is allowed to predict choice within their own equation (i.e. on for DCE, one for TT, and one for ESCAPE), or each value is allowed to predict choice in its own unique equation. [↑](#footnote-ref-5)