

EFFECTS OF EXPERIENCE-BASED FREQUENCY LEARNING ON POSTERIOR
PROBABILITY JUDGMENTS

by

Bonnie A. Armstrong, Bachelor of Arts,
University of Waterloo, 2007

A thesis

presented to Ryerson University

in partial fulfillment of the
requirements for the degree of

Master of Arts

in the Program of

Psychology

Toronto, Ontario, Canada, 2015

© (Bonnie A. Armstrong) 2015

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means. In total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Abstract

Effects of Experience-Based Frequency Learning on Posterior Probability Judgments

Master of Arts, 2015

Bonnie A. Armstrong

Psychology

Ryerson University

Aging is associated with an increase in the frequency of medical screening tests. Bayesian inference is used to estimate posterior probabilities of medical tests such as positive or negative predictive values (PPVs or NPVs). Both laypeople and experts are typically poor at estimating PPVs and NPVs when relevant probabilities are communicated descriptively. Decision making research has revealed dissociations between described and experience-based judgments. This study examined the accuracy of posterior probability estimates of 80 younger and 81 older adults when statistical information was presented through description or experience. Results show that both younger and older adults can make more accurate posterior probability estimates if they experience probabilities compared to when probabilities are described as either natural frequencies or conditional probabilities. Results also indicate that most people prefer to rely on physicians to make their medical decisions regardless of how confident they are in their judgments of probabilities.

Acknowledgements

I would like to deeply thank my supervisor Dr. Julia Spaniol for her guidance throughout this project. Dr. Spaniol's expertise, understanding and infinite patience gave me the opportunity to study a topic of great interest to me. It was a pleasure to work with her. I would also like to thank my committee members, Dr. Lixia Yang and Dr. Maureen Reed for their thoughtful suggestions. I must also acknowledge Dr. Peter Wegier for his suggestions and general interest in the project. Dr. Wegier spent much time thinking about the results and what they could mean in both a theoretical and practical sense. I would also like to thank my family for the infinite support they provided me throughout the process. Specifically, I must acknowledge Tim Rossy. His support and encouragement is appreciated, and the time he set aside to discuss research-related ideas made the process even more enjoyable. A sincere thank you to you all.

Table of Contents

CHAPTER 1: INTRODUCTION AND THEORETICAL BACKGROUND.....	1
1.1 SHARED, PATIENT-CENTERED MODEL	3
1.2 HEALTH STATISTICS	4
1.3 PROBABILITIES IN HEALTH RISK COMMUNICATION.....	5
<i>Bayesian statistics.</i>	6
<i>Physician-patient communication.</i>	9
<i>Probabilities vs. Frequencies.</i>	10
1.4 AGE-RELATED COGNITIVE DECLINE	12
1.5 AGING AND NUMERACY	14
1.6 DESCRIPTION-EXPERIENCE GAP	15
1.7 EXPERIENCING FREQUENCIES	16
1.8 SUMMARY.....	18
CHAPTER 2: THE CURRENT STUDY	20
2.1 METHOD	23
<i>Participants.</i>	23
<i>Design.</i>	23
<i>Background Measures and Cognitive Tests.</i>	23
MMSE.....	23
Numeracy Scale.	24
DOSPRT.....	24
PANAS.	25
DASS21.	25
Digit Symbol.....	25
Self-Assessment.....	25
<i>Stimuli and Apparatus.</i>	26
<i>Procedure</i>	26
2.2 RESULTS	29
<i>Data Analysis Overview.</i>	29
<i>Background Measures.</i>	29
<i>Estimates of Disease and Test Properties.</i>	30
Disease prevalence.....	30
Sensitivity of the test.....	30
Specificity of the test.	30
False-alarm rate of the test.....	31
<i>Posterior Probability Estimates</i>	31
PPVs of disease.....	31
NPVs of disease.	32
<i>Numeracy Groups.</i>	32
<i>Under – and Overweighting PPVs.</i>	33
<i>Frequency vs. Probability Response Type</i>	33
<i>Background Measures</i>	34
Self-assessment questionnaire	34

Confidence with Numbers	34
Task Difficulty.....	34
Answers Are Close to Correct Answers.....	34
Shared Medical Decision Making.....	35
CHAPTER 3: DISCUSSION AND SUMMARY	36
3.1 PPV JUDGMENT ACCURACY	36
3.2 DESCRIPTION-EXPERIENCE GAP	38
3.3 NPV	42
3.4 PRIOR PROBABILITIES	43
3.5 SELF-ASSESSMENT	45
3.6 LIMITATIONS	46
3.7 FUTURE DIRECTIONS	48
3.8 CONCLUSIONS	49
TABLES.....	50
FIGURES.....	60
APPENDICES	72
REFERENCES.....	99

List of Tables

Table 1: Relevant Properties of a Bayesian Problem Presented Descriptively.....	49
Table 2: Properties of Medical Diagnostic Tests.....	50
Table 3: Sample Characteristics.....	51
Table 4: Background Measures and Cognitive Tests.....	52
Table 5: Mean Absolute Judgment Error and Standard Deviations of Prevalence Estimates.....	53
Table 6: Mean Absolute Judgment Error and Standard Deviations of Sensitivity Estimates.....	54
Table 7: Mean Absolute Judgment Error and Standard Deviations of Specificity Estimates.....	55
Table 8: Mean Absolute Judgment Error and Standard Deviations of False-Alarm Estimates...	56
Table 9: Mean Absolute Judgment Error and Standard Deviations of Positive Predictive Value Estimates.....	57
Table 10: Mean Absolute Judgment Error and Standard Deviations of Negative Predictive Value Estimates.....	58

List of Figures

Figure 1: Schematic of simulation in experience condition.....	59
Figure 2. Estimation accuracy of prevalence.....	60
Figure 3. Estimation accuracy of sensitivity.....	61
Figure 4. Estimation accuracy of specificity.....	62
Figure 5. Bar graph illustrating age by format interaction for false-alarm estimates	63
Figure 6. Estimation accuracy of false-alarms.....	64
Figure 7. Estimation accuracy of PPV.....	65
Figure 8. Estimation accuracy of NPV.....	66
Figure 9. Accuracy of PPV estimates as a function of response type.....	67
Figure 10. Accuracy of PPV estimates as a function of response types for older adults.....	68
Figure 11. Accuracy of PPV estimates as a function of response types for younger adults.....	69
Figure 12. Accuracy of PPV estimates in experience and description conditions.....	70

List of Appendices

Appendix A: Description Medical Screening Test: Hilioplasmosis.....	71
Appendix B: Description Medical Screening Test: Mytoplasia.....	72
Appendix C: Description Medical Screening Test: Polykronisia.....	73
Appendix D: Description Medical Screening Test: Zymbosis.....	74
Appendix E: Version 1: Estimation Booklet.....	75
Appendix F: Version 2: Estimation Booklet.....	80
Appendix G: Mini-Mental Status Examination.....	85
Appendix H: Numeracy Questionnaire.....	87
Appendix I: Domain Specific Risk-Taking Questionnaire.....	89
Appendix J: Depression, Anxiety and Stress Scale.....	92
Appendix K: Positive and Negative Affect Scale.....	94
Appendix L: Digit Coding.....	95
Appendix M: Self-Assessment Questionnaire.....	96

Chapter 1: Introduction and Theoretical Background

Health care providers are now encouraging patients to become more involved in the decision making process regarding their medical care (Reyna, Nelson, Han, & Diekman, 2009; O'Connor et al., 2007, Simon, Loh, & Harter, 2007; Sheridan, Harris, & Wolf, 2004). Patients are being exposed to complex, quantitative medical information and are responsible for comprehending and using the information to make fully informed decisions (Reyna et al., 2009). Typically people have great difficulty comprehending quantitative information and tend to make inaccurate judgments about health statistics (Paulos, 1988; Reyna et al., 2009).

With age, the risk for disease increases and exposure to medical screening tests become more prevalent (Hasher & Zacks, 1988). Aging is associated with general and specific cognitive declines such as slowing of processing speed and limited working memory capacity (Park, 2000). Evidently, understanding health statistics is difficult for those who are cognitively intact. It seems logical to assume that cognitive decline may further compromise numeracy, problem solving and decision making skills, rendering older adults ill-equipped to make fully informed medical decisions.

Typically medical information is presented to patients descriptively (Zhu & Gigerenzer, 2006). Behavioural decision research has identified a distinct gap in decisions made when information is described relative to when information is experienced (Hertwig & Erev, 2009). Decisions made from description are typically based on a dense summary of information such as a statistical summary where probabilities are explicitly described (Hertwig & Erev, 2009). Additionally, prior research has shown that statistical information described in frequency format (e.g., out of 180 coin tosses, 90 landed heads) is easier to comprehend than when probabilities

are described (e.g., 50% of the coin tosses landed heads; Galesic, Gigerenzer & Straubinger, 2009) because frequencies show counts of occurrences that provides base rates, simplifying mental computations (Cosmides & Tooby, 1996). However, even when probabilities are represented as frequencies, adequate comprehension of medical statistics is still not widely achieved. Alternatively, decisions made from experience are based on a continuous collection of information in order of occurrence, that is, experiencing one event at a time and over a period of time (Hertwig & Erev, 2009). This requires consistent updating of experienced frequencies.

Research shows that people are able to make quite accurate judgments of probabilistic information if the information is experienced through frequencies (Cosmides & Tooby, 1996; Hogarth & Soyer, 2010). Experiencing frequencies has been suggested to rely on an automatic cognitive mechanism that is preserved with age (Hasher & Zacks, 1979). Because research has shown that decisions can be dependent on the way information is presented (e.g., description-based or experience-based) this could influence critical decisions regarding one's health care.

Within the last decade an abundance of research has begun testing different ways to present medical statistics to patients such as visual aids or interactive databases where icon arrays or graphs represent abstract quantitative information such as uncertainty in risk. However, even with visual or interactive aids most people do not reach the level of competency appropriate to make an informed decision (Garcia-Retamero & Galesic, 2010; O'Connor, Llewellyn-Thomas & Flood, 2004). It is not completely clear as to why people continue to have trouble with statistics even with aids. However, the continued difficulty people have with this task suggests that the human mind may not have evolved to problem solve or reason in this way.

In consideration to the current problem regarding patients' apparent lack of statistical competency and in light of the literature summarized above, the objective of the current thesis

was to examine whether experiencing frequency information would increase accuracy of probability estimates relative to estimates made when statistics are described. Specifically, the study tested whether laypeople of young and old age are able to make more accurate Bayesian inferences (i.e., posterior probability judgments) by experiencing frequency information through simulation compared to when probabilities are described through a typical dense, statistical summary. The study also aimed to discover whether an increase in accuracy will enhance independent decision making, such as a preference to rely on oneself to make final decisions about medical care rather than a doctor.

1.1 Shared, Patient-Centered Model

Both the Canadian and American health care systems have changed the structure of the doctor-patient relationship. Medical decision making has been shifted from a provider-centered model, in which doctors have the responsibility of making patient decisions, to a more shared, patient-centered model (Reyna, Nelson, Han, & Diekman, 2009; O'Connor et al., 2007, Simon, Loh, & Harter, 2007; Sheridan, Harris, & Wolf, 2004). Health care providers are exposing patients to evidence from research findings to support medical practices being used (Reyna et al., 2009) and are using health-related websites and pharmaceutical pamphlets as aids to communicate medical information to their patient. Medical professionals now have the responsibility of conveying risks and health implications correctly and clearly to patients for them to make fully informed medical decisions (Reyna & Brainerd, 2007). This model shift requires patients to comprehend qualitative and quantitative health information, placing the responsibility on the patient to make fully informed decisions about their medical care (Reyna et al., 2009). Many laboratory-based studies have tested healthy younger and older adults on their reasoning and problem solving skills of health statistics described through medical screening

tests and have found that people have great difficulty understanding quantitative medical information such as uncertainty in risk (Reyna & Brainerd, 2007). Specifically, Reyna and Brainerd (2007) showed that people have difficulty solving simple ratio and decimal problems, concepts that are required for understanding health-relevant risk communications.

Additionally, research conducted in the medical domain often makes the assumption that quantitative information such as uncertainty in risk is comprehended, however it is the decision made by the patient that is the true “primary dependent variable” or the most critical outcome (Schwartz, Woloshin, & Welch, 2005). One study showed that some participants used the words “probably” and “probability” interchangeably when making inferences about the probabilistic uncertainties of a disease (Schapiro, et al., 2008). “Probably” describes that an event is most likely to occur, “probability” is a mathematical term expressing the degree of likelihood of an event occurring (Schapiro, et al., 2008). Evidence from this study suggests that a portion of people do not comprehend the concept of probability (Schapiro, et al., 2008). Shared medical decision making, with its division of labor between physician and patient may not resolve in optimal decision making because some patients may not comprehend the information they are supposed to use to base their decisions on. Understanding probabilities and being able to interpret and use them to make health decisions is imperative and is the basis of health literacy.

1.2 Health Statistics

Health literacy is the ability to process and comprehend basic health information one needs to make appropriate health care decisions (Institute of Medicine, 2004). To understand medical information, one must have basic reading and writing abilities, be able to use and interpret quantitative information, and have the ability to articulate one’s thoughts effectively and clearly (U.S. Department of Health and Human Services, 2000; Institute of Medicine, 2004).

Together, these skills constitute health literacy.

People are exposed to health information every day. Daily exposure to health statistics ranges from reading the back of a cereal box with information on how to reduce levels of cholesterol, to information on the side effects of a medication about the probabilities of discomfort, to communicating with an oncologist about the effectiveness of cancer treatments and the corresponding survival rates (Reyna et al., 2009). Usually health information provides evidence of risks and benefits from prior research to inform people so they have relevant information to make optimal medical decisions for themselves.

Most research findings that describe health implications are presented to patients in numerical format (Reyna & Brainerd, 2007). The ability to understand and interpret quantitative health information is a subcomponent of health literacy and is described as health numeracy (Reyna et al., 2009). An important skill required to assess numerical health information is the ability to judge relative and absolute risks and benefits (Reyna et al., 2009). Accurately judging this information requires one to assess the magnitude of the risk, compare various options that contain different levels of risk and understand decimals, fractions, percentages, probabilities and frequencies (Reyna et al., 2009). Interpreting health statistics and using this information to optimize one's health care such as correctly understanding and following directions on a prescription label of a new medication (Reyna et al., 2009) require the most basic skills when making informed medical decisions.

1.3 Probabilities in Health Risk Communication

Prior research investigating quantitative skills suggests that people are significantly different in regards to numeric ability and that many people are innumerate (Paulos, 1988; Lipkus, Samsa, & Rimmer, 2001; Peters, Vastfjall, Slovic, Mertz, Mazzocco, & Dickert, 2006;

Woloshin, Schwartz, Black & Welch, 1999). In 2003 for example, the International Adult Literacy and Skills Survey (IALSS) measured adult numeracy on a scale from one to five (Statistics Canada, 2003). Level 1 represented the lowest ability and Level 5 advanced ability. In Canada, roughly 45% of the population 16 years of age and older attained numeracy scores at or above Level 3. Moreover, the majority of individuals who demonstrated basic mathematics could only successfully complete one-step arithmetic when all relevant information was explicitly provided and could not successfully problem solve through multi-step arithmetic tasks (Statistics Canada, 2003). This indicates that over half of Canadian adults failed to demonstrate the most rudimentary of quantitative skills or basic mathematical knowledge associated with functioning well in Canadian society (Rothman, Montori, Cherrington, Pigone, 2008). Quantitative information in particular is difficult for both the patient and the health care provider to comprehend (Reyna et al., 2009). These statistics suggest that a significant portion of people would have great difficulty comprehending basic medical statistics and may not possess the skills necessary to make informed decisions about their medical care.

Bayesian statistics. A more advanced skill commonly required in the interpretation of medical statistics is the ability to problem solve and interpret conditional probabilities. A conditional probability is the chance that an event (A) will occur, given that another event (B) has already occurred (Bernardo, 2003). Bayes' theorem provides an approach for calculating conditional probabilities and is used in medical diagnosis to estimate health risks. More specifically, Bayesian inference is used to estimate posterior probabilities and is defined as:

$$P(H|E) = \frac{P(H) * P(E|H)}{P(E)}$$

As Bayes' formula illustrates above, the prior probability is multiplied by the likelihood of the

evidence “E” if the Hypothesis “H” is true. This is then divided by the priori probability that the evidence itself is true which will result in a posterior probability value of “H” given the evidence (Bernardo, 2003). A positive predictive value (PPV) such as, the probability that a person has a disease given a positive test result, and a negative predictive value (NPV) such as, the probability that a person does not have a disease given a negative test result, are examples of posterior probabilities (Bernardo, 2003).

The following is a concrete example of a Bayesian problem that describes a breast cancer screening in a certain region developed by Gigerenzer & colleagues (2008):

- The probability that a woman has breast cancer is 1% (prevalence)
- If a woman has breast cancer, the probability that she tests positive is 90% (sensitivity)
- If a woman does not have breast cancer, the probability that she nevertheless tests positive is 9% (false-positive rate)

If a woman tests positive for breast cancer, what is the probability that she truly has the disease?

This question requires an estimate of the PPV.

To accurately estimate the posterior probability of a disease, medical information such as the prevalence of disease and the specificity and sensitivity of the diagnostic test must be available (Galesic et al., 2009). The prevalence of a disease is the probability of a population that has been found to have a disease. The prevalence is derived by comparing the estimated number of people that may have a disease with the total number of people of a population. The sensitivity of a test indicates how likely a positive test result will occur if a person truly has a disease. The specificity indicates how likely a negative test result will occur if a person truly does not have a disease (i.e., a true negative rate). Its inverse, the false-alarm rate, is the probability of an

inaccurate diagnostic test result, such as the likelihood of receiving a positive test result when one does not have a disease, or the likelihood of receiving a negative test result when one actually does have the disease. See Table 1 for an example developed by Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, and Woloshin (2008).

Prior research has shown that most people have such difficulty making Bayesian inferences that some researchers suggest the human mind did not evolve to process or reason through problems in this way (Zhu & Gigerenzer, 2006; Cosmides & Tooby, 1996; Kahneman & Tversky, 1972). Between 2006 and 2007, Gigerenzer and colleagues trained 1,000 gynecologists in risk communication as part of their continuing education (Gigerenzer et al., 2008). At the beginning of the first session of the continuing education course the researchers provided 160 gynecologists with enough information to accurately estimate the chance of a woman having a disease given a positive test result; that is, the participants were asked to estimate the PPV. The researchers also provided four multiple choice answers that included the correct answer and three falsely alarmed answers and a blank space. The gynecologists could either choose their answer from the multiple choice answers provided or they could generate the answer themselves using the blank space and demonstrate skills one should have as a gynecologist. The results showed the majority of gynecologists vastly overestimated the probability of cancer. Answers also showed high variability ranging from estimates of 1% to 90% chance of cancer. The number of health care professionals to answer correctly was slightly less than chance.

Another study conducted by Reyna & Brainerd (2007) tested physicians' and health care providers' knowledge of patients' diagnoses and test results specifically examining posterior probability estimates. Similar to the results from the study above, the results obtained by Reyna & Brainerd (2007) also show that highly educated people have great difficulty estimating the

PPV of having a disease, even when the type of medical information being assessed is in the physician's area of expertise (Reyna & Brainerd, 2007). These findings indicate that even trained health care professionals struggle in the estimation and interpretation of health statistics like probabilities, proportions, and risks.

Physician-patient communication. Many patients place a great deal of trust in their physician's knowledge regarding medical information, not considering that the physician could have great difficulty understanding statistics themselves and miscommunicate medical information. Without a complete understanding of health statistics, erroneous judgments could be made regarding the side effects and success rates of different treatments, which could induce unnecessary stress in the patient and could directly affect medical decision making and have serious consequences for quality of care.

Gigerenzer et al. (2008) describes the relationship between the physician and the patient to be based on the patients' automatic reliance and trust in the physician's knowledge and judgment. Yet there is an abundance of research that describes physicians' and patients' collective innumeracy, otherwise termed "collective statistical illiteracy", and is a largely unknown problem in health care (Gigerenzer et al., 2008). Health care providers need to comprehend at least basic statistics to fully inform their patients about their medical care. If health care providers lack necessary statistical skills they may heavily rely on aids such as advice from colleagues or pharmaceutical pamphlets instead of utilizing medical information based on individual patient history and current health status, and problem solving without relying too much on aids (Gigerenzer et al, 2008). This research supports the theoretical notion that people have great difficulty with numerical processing and interpretation, regardless of education level. Numerical problem solving is unavoidable, especially in the domain of medicine. This causes a

major problem for over half of the adult Canadian population who have shown to be low in numeracy (Statistics Canada, 2003).

Probabilities vs. Frequencies. Prior research has tested whether presenting statistical information in different formats would improve estimate accuracy and interpretation of posterior probabilities. Research by Galesic et al. (2009) tested participants' ability to draw Bayesian inferences when information was presented descriptively and represented as conditional probabilities and natural frequencies. The conditional probability format of risk used in Bayesian statistics is the probability of an event occurring given that another event has already occurred (Feinberg, 2006). For example, the probability that a woman actually has breast cancer given a positive test result is an estimate of a conditional probability. Put simply, it is the probability for a single event to occur (Cosmides & Tooby, 1996). This is a multistep process of numerical interpretation that requires intellectual abstraction where only relevant probabilities should be defined and used to problem solve. Not surprisingly, people tend to have great difficulty estimating and comprehending conditional probabilities (Gigerenzer et al., 2008).

The relative frequency format of risk used in Bayesian statistics is the absolute frequency divided by the total number of events (Cosmides & Tooby, 1996). For example, the amount of women testing positive for breast cancer and actually having breast cancer involves estimating a relative frequency. It requires estimation of multiple occurring events. Frequencies are the input and output of Baye's formula (Cosmides & Tooby, 1996). Gigerenzer et al. (2008) suggest that relative frequencies facilitate computation because they reduce necessary calculations and represent the way humans have cognitively developed to encode information before probabilities were invented in the mid-17th century and more abstract mathematical concepts were introduced.

Research has shown that people tend to make more accurate Bayesian inferences when

risks are presented in a natural frequency format compared to the multistep conditional probability format (Cosmides & Tooby, 1996; Reyna & Brainerd, 2007; Galesic et al., 2009). A natural frequency represents the count of occurrences, which preserves the base rate, and the representative population for the occurrences to happen (Galesic et al., 2009). The use of a base rate is critical for comparing probabilistic occurrences or prior probabilities. For example, 1% of a population are physicians and 99% of the population are not physicians, the base rate of physicians within this population is 1% (Gigerenzer et al., 2008). Galesic and colleagues (2009) suggest that natural frequencies are easier to estimate because they preserve base rates, allowing more simplified mental computations. Zhu and Gigerenzer (2006) show strong evidence in support of this theory by showing that even children can accurately solve Bayesian problems when information is represented in a natural frequency format.

However, presenting probabilistic information in frequency format does not always resolve people's difficulty in making accurate probabilistic inferences. Some research suggests that people continue to have difficulty comprehending natural frequencies, especially when faced with medical information (Garcia-Retamero, Galesic, & Gigerenzer, 2010). Researchers have found that people can be prone to biases when interpreting simple statistics. Specifically people tend to pay too much attention to numerators (i.e., the number of times an event occurred) and pay too little attention to denominators (i.e., the overall number of opportunities for the event to happen; Garcia-Retamero et al., 2010). Denominator neglect, so termed by Reyna and Brainerd (2008), can alter the meaning of a statistic significantly. For example, if 10 patients died from surgery, the level of risk would depend on the overall number of patients who had the surgery. If the overall population who underwent this procedure was 15, the surgery would pose a much bigger risk than if 1000 patients underwent the procedure.

Taken together, more accurate Bayesian inferences are made when relevant information is presented in a natural frequency format compared to when information is presented in a conditional probability format. However, even when numerical information is presented as frequencies, people continue to make inaccurate Bayesian judgments.

1.4 Age-Related Cognitive Decline

Older adults may be exposed and required to assess health statistics more often than others due to an increased risk of disease with advancing age. Additionally, older adults may have difficulty understanding health-related statistics because of age-related cognitive decline. Cognitive decline is a normative aspect of aging that can negatively impact quality of life in older adults (Hasher & Zacks, 1988; Deary et al., 2009). This poses a significant challenge in countries such as Canada, where older adults aged 65 and older constitute the fastest-growing segment of society (Statistics Canada, 2011).

Analyzing, judging and making decisions based on medical information requires multiple cognitive processes. Patients making medical decisions must be able to remember a variety of available options, rank them by importance while considering the risks and benefits, and make a final decision based on these options while contemplating future outcomes (Liu & Gonzalez, 2007). The age-related decline in cognitive capacity has been associated with forgetfulness, a decrease in the ability to maintain focus, and a decrease in the ability to reason and problem-solve (Liu & Gonzalez, 2007). Research shows that these abilities are associated to comprehending and using quantitative information to problem solve (Liu & Gonzalez, 2007).

There is a substantial amount of research investigating how cognitive aging affects older adults' comprehension and remembering of medical information (Brown & Park, 2002; Halter, 1999; Park & Kidder, 1996). This prior research has shown that older adults' judgment and

understanding of medical information is negatively affected by the aging process specifically in regards to fluid cognitive mechanisms such as working memory and processing speed (Park, Lautenschlager, Hedden, Davidson, Smith, & Smith, 2002). A decline in working memory can lead to great difficulty in understanding lengthy or complex information, specifically affecting older adults' ability to inhibit irrelevant information (Hasher & Zacks, 1988). Slowing in processing speed is described as a decline in the ability to complete a task successfully under a time constraint (Park, 2000). Moreover, older adults have shown evidence of slowing in processing speed when given a novel task or feel pressure to complete a task successfully under a time limit compared to younger adults (Park, 2000).

Age-related changes in brain structure and function may also have effects on judgment and decision making (Peters, Finucane, MacGregor, & Slovic, 2000). Functional neuroimaging techniques show a decrease in volume in the frontal cortex (Raz, 2000) and an increase in bilateral recruitment of brain regions with age (Cabeza & Dennis, 2013). Results from earlier studies suggest that older adults require recruitment of additional brain regions to compensate for age-related frontal lobe shrinkage or decrease in neural circuitry (Raz, Briggs, Marks, & Acker, 1999; Reuter-Lorenz, Marshuetz, Jonides, & Smith, 2001). These neurological changes may affect cognitive processes required for comprehending and using quantitative information.

Prior research has also revealed a linkage between the prefrontal areas and the frontal lobes, and the decline of executive functioning with age (Alvarez & Emory, 2006). Executive functioning refers to the management and regulation of cognitive processes in conscious awareness (Elliott, 2003) such as working memory, reasoning, task flexibility, problem solving, planning and execution (Monsell, 2003). If executive functioning deteriorates with age, this may be one reason older adults find comprehending and using quantitative information such a

difficult task.

If working memory, processing speed, and executive functioning tend to decline with age, along with changes in the structure and volume of the brain, this collective decline may negatively affect cognitive processes required to understand and interpret statistical information. Making health-related decisions challenges working memory and cognitive processing ability by requiring individuals to comprehend and compare multiple options, often in a stressful environment. The age-related change in frontal function may slow cognitive processing and working memory capacity, abilities that are crucial for problem solving and making subsequent judgments for optimal decisions.

1.5 Aging and Numeracy

In 2003, results of the IALSS showed 66% of the population aged 56-65 scored lower than average on the adult numeracy scale (Statistics Canada, 2003). Results also showed that 88% of adults aged 66 and older scored lower than average on the adult numeracy scale (Statistics Canada, 2003). Statistical literacy requires “online-processing” and reasoning of numeric information with dependence on sufficient cognitive resources to consciously problem solve. Probabilistic reasoning relies heavily on explicit memory. An abundance of research has shown that explicit memory deteriorates with age, whereas implicit memory tends to be preserved (Graf, & Ryan, 1990). Explicit memory is the conscious, intentional recollection of previous experiences and information, whereas implicit memory uses past experiences to remember things without conscious awareness (Graf & Ryan, 1998). The age-related deterioration of explicit memory could be one of the main causes older adults show such difficulty with numeric reasoning. Perhaps an intact explicit memory is an important component in understanding and using numerical information to inform decisions.

The problem identified in the current thesis is the responsibility older adults have to make decisions about their medical care when they are ill-equipped to do so. Unfortunately, the change in health care practice to a patient-centered model has the potential to negatively affect the health care of older adults. This model shift of medical decision making requires older adults and other vulnerable populations to make decisions that not only require high levels of numeracy and reasoning but could be life threatening in some cases.

The importance of increasing health literacy amongst the general population has been called an ethical imperative for health care providers (Gazmararian, Curran, Parker, Bernhardt, & DeBuono, 2005; Woloshin and Schwartz, 2002). Medical professionals have the responsibility to ensure that patients process and understand health-related information before decisions are made. If few doctors are able to judge the posterior probabilities of a disease, it seems unreasonable to expect untrained laypersons experiencing cognitive decline to problem solve and comprehend complex numerical information to make fully informed decisions about their medical care. More research is needed to investigate how to communicate medical information to vulnerable populations such as older adults and use the information to make optimal decisions for themselves.

1.6 Description-Experience Gap

Prior research has shown that choices can be dependent on how information is presented (Hertwig & Erev, 2009). Information that is described relative to information that is experienced can have different effects on how people interpret information and subsequent decisions that are made. Specifically, research shows that rare events tend to be overweighted when information is described and underweighed when information is experienced (Hertwig, Barron, Weber & Erev, 2004) demonstrating a reversal of preferred choice, otherwise termed the *description-experience*

gap (Hertwig & Erev, 2009). Rare events may be underweighed when information is experienced because they are not experienced often and therefore may not be remembered as well as common events that are experienced often (Hogarth & Soyer, 2014). However, when “conditional simulations” are experienced such as a Bayesian updating problem, rare events tend to be underweighed much less (Hogarth & Soyer, 2014). A simulation is a way of demonstrating random events such that observing simulated outcomes resembles real-world outcomes (Hogarth & Soyer 2014). Prior research has shown that observing simulated outcomes can facilitate comprehension of complex mathematical concepts. For example, “if the simulation is built for incidents conditional on a certain prior event, then the rare outcomes would be more visible (e.g., the probability of having a disease, given a positive result in a medical test; or the possibility of a loss beyond a certain amount, given the occurrence of a natural disaster)” (Hogarth & Soyer, 2014, p. 5). Regarding the problem identified in the current thesis, if individuals have great difficulty making Bayesian inferences when information is described explicitly, I predict that observing simulated outcomes will enhance Bayesian performance in both younger and older adults and facilitate comprehension of medical statistics.

1.7 Experiencing Frequencies

Hasher and Zacks (1979; 1984) and Zacks and Hasher (2002) have shown that humans are sensitive to regularities in their environment and are able to encode frequency information automatically. It is hypothesized that this ability is a basic cognitive mechanism that was most likely developed through evolutionary pressure (Hogarth & Soyer, 2010). To further support this claim, evidence from non-human species have shown similar patterns of encoding, e.g., animals show understanding of frequency distribution through foraging for food and greater chance of reproduction (Hogarth & Soyer, 2010). Gigerenzer (1991) has hypothesized that the human mind

is an effective “intuitive statistician” because it is able to represent probabilities as frequencies effortlessly.

Frequencies are probabilities or chance events presented in simplified form. Experiencing frequencies allows one to establish a better sense of how the data or information is proportioned. Instead of reasoning and problem solving probabilities comprehensively such as reasoning through this probability – *how likely is it that one has a disease, given a positive test result*, experiencing frequencies breaks down the information into independent events allowing one to become familiar, develop an intuition of the information and characterize proportions. Encoding frequency information allows one to experience information sequentially –independently of all other data, and over a period of time or space (Hogarth & Soyer, 2010). This activity requires little attention and does not require intention or conscious effort (Hogarth & Soyer, 2010). Hogarth and Soyer (2010) conducted a study that was designed to exploit the human natural ability to encode sequentially observed frequency information. Specifically they examined the effects of experiencing sequentially simulated outcomes and inferences made on the probabilistic implications of a regression model of an economic investment. The task involved participants experiencing the simulation of a coin toss by using a mouse to click on a computer screen and subsequently answering questions regarding the outcomes of the coin toss. The majority of participants provided answers that felt intuitive to them and were remarkably accurate, regardless of statistical sophistication that varied across participants. The authors of this study concluded that people are able to perform difficult and complex probabilistic tasks if the information is experienced sequentially.

Tversky and Kahneman (1974) proposed that the human mind did not evolve to learn complex statistical rules like probability because we do not experience them in daily life.

Perhaps human cognition does not embody a calculus of probability and this is why we have such difficulty reasoning in this way (Cosmides & Tooby, 1996). As previously discussed, physicians are taught statistics so they are able to correctly judge and estimate diagnostic test results. If highly educated people continue to have difficulty understanding probabilities and inaccurately judge risks even after years of training, it seems fair to entertain the idea that the human mind did not develop to problem solve this way (Cosmides & Tooby, 1996).

Cognitive psychologists aim to discover, among other things, what underlying cognitive mechanisms are features of the human mind for information processing (Cosmides & Tooby, 1996). The ability to automatically encode frequency data could be a factor of natural selection. Frequency performance has been shown to be unaffected by competing task demands, the amount one practices the task, or the accuracy of one's test expectations (Cosmides & Tooby, 1996). Hasher, Zacks and colleagues (1979; 1984) found no reliable trends of individual differences, and children are just as successful at encoding frequency stimuli as young and older adults are. This evidence strengthens the idea that encoding frequency information is an implicit, automatic cognitive mechanism humans naturally develop, and is preserved with age.

1.8 Summary

Aging is associated with an increase in the frequency of medical screening tests (e.g., cancer screening). Interpreting the results of such tests requires Bayesian inference. In particular, the probability of having a disease given a positive test result (PPV) must be inferred from a set of prior probabilities: the prevalence of the disease, the sensitivity and false-alarm rate of the test. Both laypeople and experts are typically poor at estimating posterior probabilities when the relevant information is communicated descriptively as probabilities, with a slight increase in Bayesian performance when information is described in frequentist terms (Gigerenzer et al.,

2008). Researchers have shown that humans are able to encode frequency information automatically and that this ability is preserved with age, supporting the hypothesis that the human mind is a good intuitive statistician (Cosmides and Tooby, 1996; Hogarth & Soyer, 2010).

At this point, however, the details of how people comprehend and interpret probabilities are not well understood, nor how the factors of format, age-related differences in abilities, and numeracy influence comprehension and interpretation. The goal of the current study was to improve our understanding of risk communication by systematically investigating how these factors influence comprehension of probabilities.

Chapter 2: The Current Study

Given the well documented gap between description-based and experience-based judgments and decisions (Hertwig & Erev, 2009), the first objective of the current study was to replicate previous research that shows inaccurate estimates of posterior probabilities when information is presented descriptively, with a specific hypothesis predicting more inaccurate estimates made when information is described through probabilities compared to frequencies by both younger and older adults. The second objective was to test whether experiencing frequency information through simulation would improve posterior probability estimates in both younger and older adults. To test this, younger and older adults completed 4 medical screening tasks in which they were randomly assigned to either a description condition or an experience condition.

In the description condition, participants read a vignette describing statistics of fictitious diseases. The purpose of this condition was to replicate previous findings by Galesic et al. (2009). Galesic et al. (2009) had participants read two vignettes about two separate diseases. The statistical information for one disease was described in a frequency format and the other disease was described in a probability format. Subsequently, participants estimated the PPV. The results of this study showed that people estimate PPVs more accurately if probability information is represented as frequencies as opposed to probabilities. However, in the current experiment 2 additional diseases were included, for a total of 4 fictitious diseases. This was done to examine whether varying the prior probabilities, and therefore posterior probabilities, would have an effect on estimate accuracy and to also examine whether response type (in terms of frequencies vs. probabilities) would affect estimate accuracy. In a further departure from Galesic et al. (2009), participants were required to estimate not only the PPV, but also the prevalence of the disease, the sensitivity, specificity, and false-alarm rate of the test, as well as the NPV. Three of

these 6 probabilities (prevalence, sensitivity and false-alarm rate) were explicitly given in the vignette. Questions about these probabilities thus served as a comprehension check: If participants understood the vignette they would be able to recognize that 3 questions asked for probabilities that were explicitly provided. The remaining 3 probabilities (i.e., the specificity, which is the inverse of the false-alarm rate, PPV and NPV) were not explicitly provided and thus had to be estimated. Table 2 shows the original problem used by Galesic et al. (2009). Problems 1-4 in the table show the properties of the disease and medical tests participants saw in the current study. The procedure was repeated for each of 4 fictitious diseases.

In the experience condition, participants experienced a simulation of a representative sample of 100 fictitious patients who had undergone the screening test. Patients were presented one at a time. Each patient was characterized by a combination of disease status (disease or no disease) and test result (positive or negative). Thus, rather than receiving explicit probabilities descriptively, participants experienced the relative frequencies of disease and test result combinations through simulation of a representative population on the computer. After experiencing the simulation, participants were asked to estimate the same probability questions given to participants in the description condition. The procedure was repeated for each of the 4 diseases.

The following hypotheses were tested:

1. In line with the idea that humans are able to encode frequency information automatically and have difficulty comprehending described probabilities, I hypothesized that both older and younger adults would provide more accurate posterior probability estimates if frequency information was experienced through simulation compared to if information is represented descriptively, either in frequency or probability format.

2. I predicted that older adults would show lower accuracy than younger adults when estimating probabilities in the description format because of the task's reliance on reasoning and problem solving abilities, which have been shown to decline with age. However, I hypothesized that both cohorts would perform similarly if frequency data are experienced through simulation, possibly due to a reliance on a preserved cognitive mechanism of encoding frequency information. If this hypothesis was supported, older adults would show more improvement between tasks compared to younger adults.
3. Numeracy level was predicted to be associated with judgment accuracy in the description condition. High numeracy requires reasoning and problem solving skills that may decline with old age (Galesic, et al., 2009). Encoding frequency information through simulation may not require numerical skills and instead may rely on an automatic cognitive mechanism that is preserved with age. That is, low and high numeracy groups in the experience condition may perform similarly. Experiencing information through simulation may help those of low numeracy perform similarly to those of high numeracy because prior literature shows that humans are typically good at encoding frequency information. However, participants who show low or high levels of numeracy may show significant differences in estimate accuracy if they are presented with information in a descriptive format because of the task's reliance on statistical literacy and numerical ability. Specifically, those that are high in numeracy may provide more accurate estimates in the description condition relative to those who are low in numeracy, due to the tasks demand of numerical skills.

2.1 Method

Participants.

All participants gave written informed consent for the study, which was approved by the Research Ethics Board at Ryerson University. Participants in the final sample included 81 younger adults and 80 older adults. Younger adults were recruited through a participant database at Ryerson University and were rewarded with 1 course credit for participating. Older adults were recruited through the Ryerson Senior Participant Pool (RSPP) and received \$12 for their participation. Four additional younger adults and 10 additional older adults were excluded for failing to meet one or more of the criteria for inclusion: absence of major health problems (e.g., history of neurological or psychiatric illness, cancer, cardiovascular disease), normal, or corrected-to-normal vision and hearing, and a score of 27 or higher on the Mini-Mental Status Examination (MMSE; Folstein et al., 1975). Characteristics of the final sample are shown in Table 3.

Design.

The design was a 2 x 2 x 4 mixed design, with age group (younger vs. older) and format (description vs. experience) as between-subject factors and disease (Hilioplasmosis, Mytoplasia, Polykronisia, Zymbosis) as a within-subject factor. Before the study began, the experimenter provided an overview of the study and participants were first asked to sign a consent form. Subsequently participants completed a computer task and then a set of background measures. Dependent variables were the estimates made for 6 probabilities (i.e., prevalence, sensitivity, specificity, false-alarm rate, PPV and NPV) for each of 4 fictitious diseases.

Background Measures and Cognitive Tests.

MMSE. Before participants completed the computer task all participants completed a

Mini-Mental State Exam (MMSE) used as a screening process to participate in the study (see Appendix G). All participants had to score 27 or above to qualify. The MMSE was developed by Folstein (Folstein, Folstein, & McHugh, 1975) and was used to measure cognitive function. It includes 19-items related to orientation, attention, memory, language and visual-spatial skills.

Once eligibility was confirmed all participants continued the study by working on the computer task. After the computer task, all participants completed a set of background measures and cognitive tests that measured numeric ability, risk propensity, working memory, emotional state and self-reflection on performance of the computer task.

Numeracy Scale. All participants completed a numeracy scale that measured general numeric ability (see Appendix H). The numeracy scale consisted of 11-items from Lipkus, Samsa, & Rimer (2001) and 1 coin toss item from Schwartz, Woloshin, Black, & Welch, (1997) for a total of 12-items and was used to measure numeric ability. Both younger and older adults had an average score of 9 on the numeracy scale, therefore participants were split and grouped into high and low numeracy groups based on whether they scored 9 and above, or 8 and below on the numeracy scale (similar to how Galesic et al., 2009 formed numeracy groups in their study). Specifically, 33 younger and 40 older adults scored 8 and below therefore representing a low numeracy group. 47 younger and 41 older adults scored 9 and above and therefore represented a high numeracy group.

DOSPERT. Next participants completed a Domain-Specific Risk Taking questionnaire (DOSPERT), developed by Weber, Blais & Betz (2002) which consisted of 42 items and measured apparent risk and risk perception in 6 domains (18 items per domain) such as ethical, financial, health and safety, social and recreational risk taking (see Appendix I). Preliminary analysis did not reveal an age or group difference, therefore this measure will not be discussed

any further.

PANAS. Participants completed a Positive and Negative Affect Schedule (PANAS) questionnaire that measured current positive and negative moods (see Appendix J). The PANAS was developed by Watson, Clark, and Tellegen (1988) and consisted of 20 self-report items (i.e., 10 positive and 10 negative items). Participants rated how much they agreed with each item on a scale from 1 (slightly or not at all) to 5 (extremely) to produce a score of their mood.

DASS21. Next participants completed the Depression, Anxiety and Stress Scale (DASS₂₁; see Appendix K). The DASS₂₁, developed by Lovibond (1995), is a 21-item questionnaire comprised of three self-report scales that was used to measure negative emotional states of depression (7-items), anxiety (7-items), and stress (7-items) that measured negative emotional state felt over the past week.

Digit Symbol. Participants continued the background measures by completing the Digit Symbol task from the Wechsler Intelligence scales (1997; see Appendix L). The purpose of the task was to correctly fill in as many symbols in the appropriate blanks that corresponded to specific digits in less than 2 minutes (i.e., all participants were given 2 minutes to complete the task). The number of symbols correctly filled in provides a measure of processing speed.

Self-Assessment. Lastly, all participants completed a self-assessment questionnaire designed specifically for the current study (see Appendix M). The self-assessment asked participants to report their confidence level and how comfortable they are working with numbers in general, how difficult the computer-task was, how close they believe their estimates they provided are to the correct answers, and who they would rely on to make final decisions about their medical care – themselves or the physician. The last page of the questionnaire provided open space for participants to descriptively comment in their own words on the strategies they

used to make estimates, how difficult they thought the task was and why, and any other comments about their experience during the experiment.

Stimuli and Apparatus.

E-Prime 2.0 (Psychology Software Tools, Inc.) was used for stimulus presentation and response collection on a 16.0" LCD display running 32-bit Windows 7 Enterprise Edition. Viewing distance was approximately 50 cm. Text instructions for the description condition appeared in black against a white background and task stimuli for the experience condition appeared in red and blue 18-point Times New Roman font against a white background.

Procedure

After providing written consent, participants completed the MMSE test. Subsequently, younger and older adults were randomly assigned to one of two conditions: the description condition that described probability information through a vignette or an experience condition where participants experienced frequency information through simulation. Both conditions allowed the participant to get a sense of how common the diseases were within a representative population and how accurate the diagnostic tests were.

In the description condition participants were shown 4 vignettes that were presented in random order to participants one at a time on the computer screen. The vignettes described the relevant statistics of fictitious diseases and their corresponding diagnostic tests (i.e., prior probabilities). The disease names were fictitious to minimize the influence of prior knowledge of diseases on task performance. For each disease, participants were given up to 10 minutes to read the vignette and answer the 6 probability questions. Vignettes for 2 of the 4 diseases, Hilioplasmosis and Mytoplasia, employed a natural frequency format (see Appendix A and B). Vignettes for the other diseases, Polykronisia and Zymbosis, utilized a conditional probability

format (see Appendix C and D). The assignment of diseases to format (probability vs. frequency) were not counterbalanced, that is Hilioplasmosis and Mytoplasia vignettes were always presented in a natural frequency format and Polykronisia and Zymbosis vignettes were always presented in a conditional probability format. When Bayesian inference problems were described as natural frequencies participants were required to answer in a natural frequency format by estimating a number as a numerator and a number as a denominator they felt to be the most appropriate estimation (see Appendix E, medical screening tests 1 and 3). When problems were described as conditional probabilities participants were required to make their estimates in percentage format (see Appendix E, medical screening tests 2 and 4).

In the description condition, the prevalence of the disease, the sensitivity and false-alarm rate of the test were explicitly provided. The accuracy of estimates made for these properties thus served as a comprehension check. Accuracy of estimates made for the remaining properties, that is, the specificity, positive and NPVs were used to measure reasoning and numerical problem solving ability. Of critical importance were participants' estimates of the positive and NPVs in analyzing how accurate people are at estimating posterior probabilities; these were the primary dependent variables. Participants' estimates were subtracted from the true values to obtain absolute deviation scores.

In the experience condition, participants experienced a simulation of a representative sample of 100 fictitious patients who had undergone a screening test (see Figure 1). Patients were presented one at a time for three seconds and centered on the computer screen. Each patient was separated by a blank white screen presented on the computer screen for 1 second. Patients were characterized by a combination of disease status (disease or no disease) and test result (positive or negative). The words "Has Disease" and "Positive Test Result" were presented in

red font and the words “Does Not Have Disease” and “Negative Test Result” were presented in blue font. The number of patients representing each of the four combinations of disease status (has disease vs. does not have disease) and test result (positive vs. negative) differed for each disease (see Table 2). Participants were instructed to not write anything down, nor attempt to use memorization, but to simply pay attention to the information on the screen. Once participants had viewed the 100 patients, their task was to estimate the 6 probabilities corresponding to properties of the disease and the medical test.

Two of the 4 populations experienced through simulation required participants to make estimates in a natural frequency format (e.g., “the false-alarm rate is 30 out of 100”). The remaining 2 representative populations required participants to estimate in a probability format (e.g., “the false-alarm rate is 30%”). Identical to the description condition, in order to analyze estimate accuracy we took the absolute difference between the participants’ estimate and the true value as a continuous measure of accuracy.

Two versions of the response booklet were created to counterbalance the response format (i.e., probability vs. frequency format). The first booklet version had participants estimate in the following order: frequency, probability, frequency, probability (see Appendix E). The second version had participants estimate in the following order: probability, frequency, probability, frequency (see Appendix F). After the computer-task of estimating probabilities was complete, participants filled out 6 background measures.

2.2 Results

Data Analysis Overview.

Performance was quantified by subtracting participants' estimates from the true values to obtain absolute deviation scores (i.e., error scores). A mixed ANOVA of age (younger, older), format (description, experience) and disease (Hilioplasmosis, Mytoplasia, Polykronisia, Zymbosis) was performed on the error scores for all 6 probabilities participants were asked to estimate. All analyses employed the Bonferroni corrections for multiple comparisons and any violations of sphericity found during analysis were corrected using the Greenhouse Geisser correction factor. The alpha level was set to .05 and all tests were two-tailed.

A univariate analysis was performed on the absolute judgment error for each probability with age group, format and numeracy level as between-subjects factors and disease as within-subject factor. This analysis was conducted to examine whether format type had an effect on probability estimates in younger and older adults and those of high and low numeric ability. Simple effects analyses were conducted by running independent-sample t-tests to test the effect age, format and numeric ability may have on probability estimates for each disease. For all of the background measures and cognitive tests, raw scores were used and a univariate analysis was run to establish whether the groups were equated on background characteristics (e.g., years of education). To investigate whether rare events, such as the PPVs, were under- or overweighted in each condition, beanplots were used to display the distribution of deviations.

Background Measures. The younger adult sample included more females than males, whereas the older adult sample included a slightly more balanced group of females and males (see Table 3). All older adults included in the sample had MMSE scores of 27 or greater ($M = 29.13$, $SD = 1.15$). For an educated sample, scores of 26 and below may indicate cognitive

impairment. Older adults had significantly more years of education compared to younger adults. However, younger adults performed more proficiently on the Digit Symbol Coding task and scored significantly higher on all DASS subscales revealing higher depression, anxiety and stress symptoms within this group (see Table 3). Older adults reported significantly greater positive affect scores relative to younger adults, whereas younger adults reported significantly greater negative affect scores than older adults (see Table 3). Finally, correlational analyses performed within each group of participants showed that background characteristics did not significantly correlate with any experimental outcomes.

Estimates of Disease and Test Properties.

Disease prevalence. There was a significant main effect of format, $F(1, 157) = 7.99, p = .005$, partial $\eta^2 = .048$, with more accurate estimates of the prevalence of diseases in the description condition ($M = 1.06$) compared to the experience condition ($M = 4.37$). There was no significant effect of disease, $F(2.67, 425) = 1.89, p > .05$, partial $\eta^2 = .012$, and no significant difference between age groups, $F(1, 157) = .077, p > .05, \eta^2 = .000$. Overall, both younger and older adults estimated the prevalence of all four diseases more accurately in the description condition compared to the experience condition (see Table 5 and Figure 2 for more detail).

Sensitivity of the test. There was a significant main effect of format, $F(1, 157) = 71.15, p < .001$, partial $\eta^2 = .312$, with more accurate sensitivity estimates of the 4 diagnostic tests corresponding to the 4 diseases in the description condition ($M = 4.48$) compared to the experience condition ($M = 32.62$). Again, there was no significant difference in estimates of sensitivity across the 4 diseases, $F(2.67, 425) = 1.34, p > .05$, partial $\eta^2 = .008$. There was also no main effect of age, $F(1, 157) = .493, p > .05$ (see Table 6 and Figure 3 for more detail).

Specificity of the test. There was a significant main effect of format, $F(1, 157) = 10.13, p$

$<.01$, partial $\eta^2 = .061$ with both younger and older adults providing more accurate estimates of specificity in the experience condition ($M = 6.36$) compared to the description condition ($M = 14.74$). There was also a significant main effect of disease, $F(2.67, 425) = 4.93$, $p < .01$, partial $\eta^2 = .03$. Specifically, participants made more accurate specificity estimates for Mytoplasia ($M = 8.22$) and Zymbosis ($M = 7.62$) and less accurate estimates for Hilioplasmosis ($M = 13.61$) and Polykronisia ($M = 12.76$). There was also a trending main effect of age, $F(1, 157) = 3.69$, $p = .056$, partial $\eta^2 = .023$, with younger adults ($M = 8.02$) providing slightly more accurate specificity estimates than older adults ($M = 13.08$; see Table 7 and Figure 4 for more detail).

False-alarm rate of the test. Results revealed a significant age by format interaction, $F(1, 157) = 7.80$, $p < .01$, partial $\eta^2 = .047$. Simple effects analysis revealed that younger adults made more accurate false-alarm estimates in the description condition ($M = 1.83$) relative to the experience condition ($M = 8.23$), $t(79) = -2.98$, $p = .004$, $d = .66$. However, there was no significant difference between older adult false-alarm estimates in the description or experience conditions, $t(78) = 1.06$, $p > .05$, $d = .23$. Independent samples t-tests were also conducted for description and experience conditions separately to compare younger and older adult false-alarm estimates between conditions. T-tests revealed a significant age difference with younger adults ($M = 1.83$) estimating more accurately than older adults ($M = 8.46$) in the description condition, $t(77) = -3.01$, $p = .004$, $d = .67$, and no significant age difference in false-alarm estimates made in the experience condition, $t(80) = .99$, $p > .05$, $d = .22$ (see Figure 5, 6 and Table 8 for more detail).

Posterior Probability Estimates.

PPVs of disease. Critically, there was a significant main effect of format, $F(1, 157) = 152.40$, $p < .001$, partial $\eta^2 = .49$, with more accurate PPV estimates made in the experience

condition ($M = 17.33$) compared to the description condition ($M = 46.47$). There was also a significant main effect of disease on the accuracy of PPV estimates, $F(2.67, 425) = 8.06$, $p < .001$, partial $\eta^2 = .049$ (see Figure 7 and Table 9 for more detail on PPV estimates as a function of condition and disease).

NPVs of disease. There was a main effect of format, $F(1, 157) = 17.62$, $p < .001$, partial $\eta^2 = .101$ with participants in the experience condition ($M = 12.49$) providing more accurate NPV estimates compared to the description condition ($M = 25.06$). There was also an age by format interaction, $F(1, 157) = 6.63$, $p < .05$, partial $\eta^2 = .04$, with older adults providing less accurate NPV estimates in the description condition ($M = 29.06$) compared to younger adults in the description condition ($M = 21.05$), $t(79) = -1.87$, $p < .01$, $d = .51$, and with older adults providing more accurate NPV estimates in the experience condition ($M = 8.79$) compared to younger adults in the experience condition ($M = 16.19$), $t(79) = -1.72$, $p < .01$, $d = .67$ (see Figure 8 and Table 10 for more detail on NPV estimates between age groups, conditions and diseases). Therefore, older adults showed a larger performance difference between conditions relative to younger adults, that is, experiencing information seemed to benefit older adults more than younger adults specifically for NPV estimates (see Figure 8.).

Numeracy Groups.

In order to analyze performance differences and the effect of numeracy on PPV estimates, a 3-way (format x numeracy group x age) between-subjects ANOVA was conducted on PPV estimates. Results revealed a main effect of format, $F(1, 157) = 21.16$, $p < .001$, partial $\eta^2 = .65$, with participants more accurately estimating PPVs in the experience condition ($M = 18.93$) compared to the description condition ($M = 52.73$). No effects involving age or numeracy group were significant.

Under – and Overweighting PPVs.

Observing the distribution of under and overweighting probabilities is important because absolute deviations are not informative about the direction of estimation errors. Figure 7 provides an illustration of participants' accuracy in PPV estimates between conditions for each disease, and the frequency of over- and underestimations. Examination of these plots suggests that participants in the experience condition estimated small PPVs (e.g., 8.33%) with high accuracy, however as the PPV increases, under- and overweighting probabilities increases. In contrast, in the description condition, participants provided more accurate PPV estimates when the true PPV was large (e.g., 50%), and less accurate PPV estimates when the true PPV was small (e.g., 8.33%) with the majority of participants typically overweighting PPV estimates.

Most participants seem to accurately estimate the NPV, with a portion of participants underestimating this value in both the experience and description conditions. This could be because the NPV has an upper limit of 100%. That is, participants can only estimate 100% or lower (i.e., a ceiling effect). Two of the 4 diseases had a NPV of 100%; therefore any deviation from the true value will typically result in underestimation (see Figure 8).

Frequency vs. Probability Response Type.

Because the response format (frequency vs. probability) was confounded with disease in the description condition, it was not possible to draw firm conclusions regarding the effect of response format on estimate accuracy. However Figure 9 displays the differences in judgment error of PPV estimates between response types from the current study. The results resemble patterns reported in previous studies that examined estimate accuracy when information is described in either a probability or frequency format (Gigerenzer & Hoffrage, 1995; Galesic, Gigerenzer & Straubinger, 2009). Participants provided less accurate PPV estimates when

relevant information was described in probability format and provided more accurate estimates when information was described in a frequency format. However, these results show that *even more* accurate posterior probability estimates are made when participants experience frequency information through simulation compared to when information is described. These patterns of results hold true for both older (see Figure 10) and younger adults (see Figure 11).

Background Measures.

Self-assessment questionnaire

Confidence with Numbers. Participants were asked how much they agree with the statement “I am confident and comfortable working with numbers”. Answers were indicated on a 5-point Likert scale with 1 representing “strongly disagree” and 5 representing “strongly agree” (see Appendix N). A univariate analysis was conducted and results revealed that younger ($M = 2.89$) and older adults ($M = 3.05$) show no significant difference in reporting how confident and comfortable they are working with numbers, $F(1, 157) = .676, p > .05$, partial $\eta^2 = .004$.

Task Difficulty. Participants were asked how much they agree with the statement “I found the task of estimating properties of a diagnostic test difficult”. Answers were indicated on a 5-point Likert scale with 1 representing strongly disagree and 5 representing strongly agree (see Appendix N). The description condition ($M = 3.70$) was rated more difficult than the experience condition ($M = 3.05$), $F(1, 157) = 18.63, p < .001$, partial $\eta^2 = .11$ with no significant effect of age.

Answers Are Close to Correct Answers. Participants were asked how much they agree with the statement “I believe the estimates I made are close to the correct answers”. Answers were indicated on a 5-point Likert scale with 1 representing strongly disagree and 5 representing strongly agree (see Appendix N). Interestingly, ratings were higher in the experience condition

($M = 3.34$) than in the description condition ($M = 3.00$), $F(1, 157) = 5.82$, $p = .017$, partial $\eta^2 = .036$. There was no main effect of age or age by format interaction.

Shared Medical Decision Making. Participants were asked the question “who do you feel most comfortable with to make final decisions about your medical care – you as the patient or the physician?” Participants were asked to circle a number from 1 (patient) to 10 (physician) to best represent who they would prefer to make final medical decisions about their health care. There was a marginal main effect of age, $F(1, 157) = 3.03$, $p = .083$, partial $\eta^2 = .019$, with younger adults ($M = 6.64$) slightly preferring to rely on a physician to make final decisions about their medical care compared to their older adult counterparts ($M = 5.93$). There was no effect of condition.

Chapter 3: Discussion and Summary

The current study examined whether experiencing a Bayesian updating problem presented as frequencies observed sequentially would improve Bayesian inferencing above and beyond inferences made when relevant information is described. Results of the current study provide evidence that both older and younger adults, and those of low and high numeric ability make more accurate Bayesian inferences when information is experienced compared to when information is described. Findings are discussed in detail below.

3.1 PPV Judgment Accuracy

Results from the current study resemble evidence found by Galesic et al. (2009). Specifically, in the description condition both younger and older adults provided more accurate PPV estimates when relevant information was described in a natural frequency format compared to a conditional probability format. These results support the literature suggesting that frequency formats provide a more transparent representation of risk compared to other formats such as probability. Contrary to results found by Galesic and colleagues (2009), evidence from the current study show no performance difference among those of low or high numeracy in the description condition, also contrary to hypothesis 3 in the current study. Although different from what Galesic et al (2009) found, the current results support prior evidence suggesting that the human mind may not embody a cognitive mechanism that facilitates probabilistic reasoning even for those who are highly numerate. However, humans may be able to process probability information more accurately if information is described as frequencies because frequencies preserve base rates, which simplify mental computation (Cosmides & Tooby, 1996).

Although participants statistically made more accurate PPV estimates when information was described through frequencies compared to probabilities, estimates made based on frequency

information still varied in accuracy. This suggests that the majority of people still lack comprehension of risk. In the real world, if probabilistic information is not comprehended or is misunderstood, this could cause major problems for those who are required to use probabilistic information to make critical decisions about their medical care. The main objective of the current study was to test whether learning probabilities through experience as opposed to description would enhance Bayesian performance. More specifically, would experiencing frequency information through a simulation further improve posterior probability estimates of a medical screening test?

The current study found that both younger and older adults are able to make more accurate posterior probability estimates if information is experienced rather than if information is described. That is, they are able to proficiently estimate complex and abstract mathematical problems that even highly educated people have difficulty calculating, if information is simulated as frequencies presented sequentially. Specifically, results showed a reduction in deviations between PPV estimates and true values by more than half between conditions. That is, participants learned about uncertainty in risk and estimated posterior probabilities quite close to true values when the information was experienced through a simulation as opposed to when the information was described in a summary.

Results also showed no difference in PPV estimates between response types (probability vs. frequency) when information was experienced. The results suggest that experiencing information facilitates judgments of PPVs regardless of answering in a frequency or probability format. Perhaps this way of learning conditional information (i.e., independent events experienced through simulation) enhances mental representation of the proportions and probabilities of the information. Observing a “conditional simulation” that demonstrates a

Bayesian problem may make probabilities more transparent and easier to understand which could then make responding in a frequency or probability format easier. The simulation allows the observer to encode frequency information sequentially, which may tap an automatic cognitive mechanism. This could be why people relate to the simulation so well. However more research is needed to understand the relationship between experiencing information and answering in different numerical formats.

The results provide additional evidence supporting the idea that the ability to learn frequency information through experience is preserved with age and that those of low numeric ability also benefit from this presentation format. These results are in line with hypothesis 3 regarding the experience condition in the current study. It has been suggested that experiencing frequency of occurrence may tap implicit memory or an automatic encoding process. Implicit memory has been well documented as a type of memory that requires no conscious effort and is preserved with age. The results suggest that if complex probabilistic information is presented in a way that depends on a cognitive mechanism that processes information automatically and does not deteriorate with age, vulnerable populations such as the elderly or those of low numeric ability will benefit from this presentation format and better equip them to make medical decisions for themselves. This adds to the growing body of literature suggesting that cognitive mechanisms that encode frequency information automatically may exist.

3.2 Description-Experience Gap

Research on the description-experience gap shows that small probabilities tend to be overweighed if information is described, however decisions from experience show that small probabilities tend to be underweighted. Regarding decisions made from experience, it has been suggested that if rare events are not experienced often people may not pay attention to them or

the events may not be salient enough to be remembered. However when conditional information is experienced (e.g., a Bayesian updating problem such as the medical screening test in the current study) rare outcomes become more salient and people tend to underweight rare events less.

In the current study true PPVs were varied to test this phenomenon. The results suggest that those in the description condition typically overestimated the PPVs and those in the experience condition did not show specific patterns of over or underweighting PPVs.

More specifically, PPV estimates made in the description condition were either vastly overestimated or underestimated relative to the true value, with more overweighing. Interestingly, further analysis revealed that as events become more rare (e.g., there is an 8.33% probability of having a disease, given a positive test result), estimates become less accurate with an increase in overweighing the value. However as the true PPVs increase (e.g., there is a 50% probability of having a disease, given a positive test result), estimates become more accurate with less overweighing and a slight increase in underweighing. These results support theories of decisions based on descriptions suggesting that rare events tend to be overweighed when information is described.

Contrary to results found for estimates made in the description condition, participants in the experience condition made more accurate PPV estimates for the diseases with the lowest true PPVs (i.e., 8.33% and 20%) and slightly less accurate PPV estimates for the diseases with higher true PPVs (i.e., 25% and 50%). Specifically, estimates were more accurate for Zymbosis (true PPV of 8.33%) and Polykronisia (true PPV of 20%). However, as the true PPV increases, such as the true PPVs for Mytoplasia (true PPV of 25%) and Hilioplasmosis (true PPV of 50%) the distribution of deviations vary more around the true value however the means remain close to the

true value. That is, the more the true PPV increases the less accurate PPV estimates become. This could be showing that as the uncertainty of a diagnostic test predicting disease increases participants' estimates become less accurate. These results suggest that the greater the uncertainty of the true PPV, the greater the possible range of outcomes, increasing the variance of estimates made, therefore decreasing one's grasp on the underlying probabilities. Further research is needed to examine why uncertainty influences inaccurate probabilistic estimates.

The current results are in line with the literature suggesting that when rare events are experienced they are underweighted, whereas when a conditional simulation is observed such as the updating Bayesian task in the current study, rare events are underweighted much less. More specifically, the results suggest that people gain an intuitive sense of events that are more rare such as a small chance of 8.33% of having a disease, given a positive test result and therefore make more accurate PPV judgments by estimating low values for a probability they sense is rare when conditional information is experienced through simulation. Estimates made by participants in the experience condition may become less accurate as the PPV increases possibly because uncertainty increases. For example, Hilioplasmosis has a true PPV of 50%. That is, there is a 50% probability that one will have the disease, given a positive test result. Zymbosis however has a true PPV of 8.33%. That is, there is an 8.33% probability that one will have the disease given a positive test result. Perhaps less accurate estimates were made for the PPV of Hilioplasmosis because of the high level of uncertainty of this probability (i.e., 50% chance one has the disease given a positive test result).

Overall, the results suggest that participants in the description condition estimate PPVs *less* accurately as diagnostic tests become less accurate and participants in the experience condition estimate PPVs *more* accurately as the diagnostic tests become less accurate. These

results provide strong supporting evidence that small probabilities tend to be overweighed if information is described and more accurate judgments of rare events are made specifically when information is experienced through simulation. The underlying PPVs used in this study essentially represent worst-case scenarios. The diagnostic tests are typically inaccurate and are unable to reliably predict if one has a disease given a positive test result, however participants in the experience condition were able to detect this unreliability of the test and provide accurate PPV estimates. The results could indicate greater comprehension of posterior probabilities when information is experienced compared to when information is described, however further research is needed to statistically test this phenomenon.

The results from the experience condition also suggest that even if a diagnostic test is extremely inaccurate and unreliable, people are able to detect the inaccuracy of the diagnostic test and comprehend that the test is unreliable if they experience the data as frequencies through simulation. Being able to accurately detect and understand how reliable a diagnostic test is will not only better inform patients of their prognosis but it could also heavily influence their judgment and decision making on their medical care. This could have a positive impact on patient-physician communication and bolster optimal decision making for a patient's condition.

Taken together, the results from this study support prior evidence suggesting that more accurate PPV estimates are made when information is described through frequencies compared to probabilities. Results also show that Bayesian performance can be further enhanced if information is experienced as frequencies simulated sequentially, regardless of age or numeracy level. These results support hypothesis 1 of the current study. Additionally, the results suggest that experiencing information through simulation can weaken the well-documented difference in performance when estimating in frequency or probability format. Observing a conditional

simulation may also minimize the tendency to underweight rare events. In addition to estimating PPVs in the current study, participants also estimated NPVs and prior probabilities.

3.3 NPV

Participants were asked to estimate the probability of not having a disease, given a negative test result (i.e., NPVs) for each disease, based on a set of prior probabilities. Typically NPVs of medical screening tests are high values as most people in a population are not diseased and most will receive negative test results. This value was tested to examine how accurate participants are at estimating common as opposed to rare events.

Younger and older adults provided quite accurate estimates of NPVs in both the description and experience conditions. This could be because the information, either read or experienced, more clearly demonstrated that there were more people in the representative populations who received negative test results and who do not have the disease compared to people who received positive test results and have the disease. In the current study, the properties of the diseases were constructed to reflect a typical population – the amount of people who received a positive test result who have the disease was generally minimal, representing a rare event and the majority of the population were healthy and received negative test results. Participants may have provided larger estimates that were close to true NPVs because they were aware that most people in the representative population were healthy and received accurate negative diagnostic test results.

However, the results revealed an age by format interaction with older adults providing less accurate NPV estimates in the description condition compared to younger adults, and older adults providing more accurate NPV estimates in the experience condition compared to younger adults. Therefore, older adults showed a greater performance difference between format types supporting hypothesis 2 of the current study. The results suggest that older adults may benefit

from experiencing frequency information through simulation more than younger adults, specifically for estimating NPVs. Older adults may have more difficulty estimating posterior probabilities when information is described compared to younger adults due to the tasks reliance on reasoning and problem solving. More specifically, the results suggest that making probabilistic inferences such as estimating the NPV based off prior probabilities may require the cognitive effort of reasoning and problem solving – skills that are compromised with age. This could be why younger adults outperformed older adults in the description condition.

Experiencing probability information and subsequently making inferences about the information may not require the cognitive effort of reasoning or problem solving, and instead may rely on a more automatic cognitive mechanism. The results support prior literature suggesting that humans are able to encode frequency information automatically and with little effort, a cognitive process that is preserved with age. However, further research is needed to examine why older adults estimated NPVs more accurately than younger adults in the experience condition.

3.4 Prior Probabilities

Participants were also required to estimate prior probabilities in addition to posterior probabilities, such as the prevalence of the disease, the sensitivity, specificity and false-alarm rate of the test. The results showed that participants provided more accurate probabilities in the description condition compared to the experience condition when information was explicitly provided in the summaries, allowing participants to provide the exact true values with no error. These results suggest that both younger and older adults are able to identify probabilities explicitly provided in statistical summaries such as the prevalence, sensitivity and false-alarm rate. However, participants typically provided more accurate estimates in the experience

condition for those probabilities that were not explicitly provided in the vignettes and problem solving was required (i.e., specificity).

Interestingly, the age by format interaction on false-alarm estimates suggest that older adults have more difficulty comprehending probabilities of receiving a false test result more than younger adults. Prior research shows that people have difficulty understanding the concept of false positive or negative test results (Reyna & Brainerd, 2007). Perhaps cognitive decline with age further compromises reasoning and understanding probabilities of false-alarm rates. Research should further investigate how people come to understand and interpret unreliable diagnostic tests and the impact cognitive decline may have.

Results from the current study also show that when calculation was required, in particular for specificity, all participants, regardless of age, benefit from experiencing frequencies through simulation over performing a basic mathematical procedure such as subtraction (i.e., $100 - \text{false-alarm rate} = \text{specificity}$). These results highlight the difficulty people have with even basic numeracy when relevant information is presented through description.

Overall, the results suggest that when probabilities are explicitly given in a summary (e.g., prevalence, sensitivity, and false-alarm rate) the majority of people are able to identify and estimate prior probabilities accurately. In comparison, results show that experiencing prior probabilities can lead to less accurate estimates. However, for probabilities that require calculation (e.g., specificity, PPV and NPV) more accurate estimates are made if information is experienced through simulation. These results suggest that basing judgments on experience make estimating more complex probabilities easier because the information is presented in a way the human mind can understand.

In summary, results critical to the objective of the current study show that both younger

and older adults made more accurate posterior probability estimates when information was experienced through simulation compared to when information was described. The results resemble prior research showing that most people inaccurately estimate posterior probabilities with a slight increase in estimate accuracy when information is described in frequency format compared to probability format. Results also show that Bayesian performance can be significantly enhanced if information is experienced allowing people to rely on a preserved, automatic cognitive mechanism that can encode frequency information with ease.

3.5 Self-Assessment

Participants also completed a self-assessment questionnaire after the computer task for reflection and to provide some insight on their performance. Results showed that both younger and older adults reported low confidence working with numbers in general, however those participants who experienced the simulation of information reported that they believe the estimates they made are close to the correct answers compared to participants in the description condition. This suggests that participants felt more confident making complex probability estimates and believed their estimates were close to the correct answers if they experienced information through simulation, even though they lack confidence working with numbers in general.

However, when asked the question “who do you feel more comfortable with to make final decisions about your medical care – you as the patient or the physician?” both younger and older adults reported that they prefer to rely more on the physician to make their medical decisions compared to them as the patient making medical decisions for themselves regardless of the condition they were randomly assigned to. Statistical tests revealed that younger adults prefer to rely on their physician more than older adults.

These results suggest that even though people feel confident in their probability estimates when they experience simulated outcomes, the confidence they have in the accuracy of their estimates may not transfer to using this knowledge to make decisions in the real world (e.g., making decisions about medical care). Results from the self-assessment questions show that participants prefer to rely on their physician to make final medical decisions even though research shows that trained professionals have just as much trouble comprehending and making inferences about abstract mathematical concepts like uncertainty in risk (e.g., posterior probabilities). Future research should further investigate the transfer of intuition and judgment of probabilities to real world situations and test different ways to apply what has been learned through simulation to make real world decisions.

3.6 Limitations

There are a number of limitations in this study and future directions to be considered. First the sample of participants tested in the study may not be representative of the population. For example the older adults who volunteered to participate are a highly educated cohort with a mean of 16.3 years of education and are motivated to contribute and learn about current, ongoing research, which could be different from older adults in the general population. Additionally, gender was extremely imbalanced in the study. Future research should include a more even male to female gender ratio.

Second, the ecological validity of the study is challenged as participants may respond differently when given a hypothetical summary of health statistics or a simulation of a representative patient population compared to a real world situation where the information is utilized to make real medical decisions. For example answering questions about fictitious diseases and the hypothetical probabilities of disease prevalence and accuracy of diagnostic tests

may not elicit the emotions associated with real world medical decision making.

Third, the PPVs chosen for the diseases range from low accuracy of a diagnostic test (e.g., there is an 8.33% probability of having Zymbosis, given a positive test result) to the accuracy of a diagnostic test predicting a disease at chance level (e.g., there is a 50% probability of having Hilioplasmosis, given a positive test result; see Figure 12). In the real world, diagnostic tests used to predict disease should be more reliable than the accuracy of the hypothetical tests used in this study (Irwig, et al., 1994). Future research should examine whether participants make more accurate estimates as the diagnostic test becomes more accurate (e.g., there is a 75% probability of having some disease, given a positive test result). In line with the idea that estimating probabilities can be easier for rare events when information is experienced, such as a small chance that a diagnostic test can predict disease given a positive test result, perhaps PPV estimates will improve as the true PPVs increase, such as a large chance that a diagnostic test can predict disease given a positive test result. Research in the future should examine whether estimate accuracy increases as PPVs become larger. If this is the case, we can draw a stronger conclusion that as the uncertainty of diagnostic tests increase, the less accurate people are in their PPV estimates even if information is experienced through simulation.

Fourth, the response format was nested within disease, resulting in a confound in the description condition. Specifically, summaries for Hilioplasmosis and Mytoplasia were always represented as natural frequencies and Polykronisia and Zymbosis were always represented as conditional probabilities. All diseases should have been described as both frequency and probability and presented randomly to participants. The results show that estimate accuracy is dependent on numerical representation (probability vs. frequency) and true PPVs. Therefore diseases with larger true PPVs were only presented as conditional probabilities and diseases with

smaller true PPVs were only presented as natural frequencies in the description condition, which could have affected the accuracy of estimates made. More research is needed to further investigate the relationship between format type (description vs. experience), numerical representation (probability vs. frequency) and varying true PPVs.

3.7 Future Directions

Future research should examine the neurocognitive mechanisms underlying learning and judgment performance in description and experience-based Bayesian inference. For example, does encoding frequency information sequentially tap implicit memory and/or explicit memory? Perhaps older adults are able to estimate just as accurately as younger adults because the task is tapping a cognitive mechanism that does not decline with age, such as implicit memory.

Future research should also consider testing the domain generality of experiencing quantitative information through simulation. Perhaps this way of presenting complex and abstract mathematical concepts will translate and enhance performance in finance as well.

Further examination of individual-difference factors for experiencing statistical information such as numeracy level, and working memory, may also help reveal information as to how people are able to comprehend and interpret information in this way.

Additionally, further research is needed to examine why there is no difference in estimate accuracy when information is represented as probabilities compared to frequencies when information is experienced, however a difference in estimate accuracy emerges between response types when information is described.

Last, a more practical direction for future research is to investigate whether enhanced Bayesian performance (presumably from learning statistical information by observing a conditional simulation as shown in the current study) translates to more informed decision

making in the real world. For example, patients may find using the simulation helpful to demonstrate complex quantitative medical information discussed by the physician. Future research should examine whether gaining a more accurate sense of the information from observing a simulation better equips patients to make more informed decisions for themselves about their medical care.

3.8 Conclusions

Patients now have the responsibility of comprehending health statistics and using this information to make critical decisions about their medical care. Research has shown that people have great difficulty performing the most rudimentary mathematics. With age, cognitive decline further compromises the necessary skills such as reasoning, cognitive flexibility and problem solving required for comprehending and interpreting quantitative information. Unfortunately, most patients are older due to the likelihood of disease increasing with age, rendering older adults ill-equipped to make informed medical decisions for themselves. The current research presents a new and exciting avenue toward helping people comprehend and interpret complex medical statistics regarding diseases and the reliability and accuracy of corresponding diagnostic tests through experiencing frequency information through simulation. Future research is encouraged to examine both the theoretical and practical implications of experiencing frequencies through simulation and the impact this has on subsequent Bayesian inference.

Tables

Table 1

Relevant Properties of a Bayesian Problem Presented Descriptively

<u>Relevant Property</u>	<u>Description</u>
Prevalence	The probability that a woman has breast cancer is 1%
Sensitivity	If a woman has breast cancer, the probability that she tests positive by a mammography screening is 90%
False-Alarm Rate	If a woman does not have breast cancer, the probability that she tests positive by mammography screening is 9%

Note. Specificity could be used as a relevant probability instead of the false-alarm rate as specificity is the inverse of the false-alarm rate.

Table 2

Properties of Medical Diagnostic Tests

<u>Properties</u>	<u>Original Problem</u>	<u>Hilioplasmosis</u>	<u>Mytoplasia</u>	<u>Polykronisia</u>	<u>Zymbosis</u>
D_Pos	9	5	3	2	1
D_Neg	1	2	1	0	0
ND_Pos	89	5	9	8	11
ND_Neg	901	88	87	90	88
Sample Size	1000	100	100	100	100
Prevalence	1.00%	7.00%	4.00%	2.00%	1.00%
Sensitivity	90.00%	71.43%	75.00%	100.00%	100.00%
Specificity	---	94.62%	90.60%	91.84%	89%
False-Alarm	8.99%	5.38%	9.38%	8.16%	11.11%
PPV	9.18%	50.00%	25.00%	20.00%	8.33%
NPV	---	97.77%	98.86%	100.00%	100.00%

Note. Table 2 abbreviations:

D_Pos: Number of patients with disease who get positive test result; D_Neg: Number of patients with disease who get negative test result; ND_Pos: Number of patients without disease who get positive test result; ND_Neg: Number of patients without disease who get negative test result.

Table 3

Sample Characteristics

	<u>Description</u>		<u>Experience</u>	
	<u>Younger</u>	<u>Older</u>	<u>Younger</u>	<u>Older</u>
N	41	40	40	40
Sex	Male (2); Female (39)	Male (15); Female (25)	Male (8); Female (32)	Male (15); Female (25)
Age (years)	20.82 (3.67)	75.13 (6.35)	20.78 (3.14)	72.44 (5.43)
Age Range	17 – 33	65 - 92	17 – 27	66 - 88
*Education (years)	14.28 (1.78)	16.36 (2.75)	14.39 (1.79)	16.20 (3.15)
MMSE	29.27 (.91)	29.18 (1.10)	29.17 (.80)	29.07 (1.21)
Numeracy	9.12 (2.15)	8.64 (2.33)	8.90 (2.38)	8.88 (2.28)
*Positive Mood	24.05 (7.65)	31.67 (6.16)	25.78 (7.81)	32.41 (6.89)
**Negative Mood	14.50 (4.75)	11.69 (4.48)	13.32 (4.10)	11.49 (2.36)
**Depression	5.65 (4.48)	2.72 (3.61)	4.71 (5.30)	1.68 (2.02)
**Anxiety	5.52 (4.18)	1.92 (1.99)	4.71 (4.37)	1.73 (2.45)
**Stress	7.52 (3.95)	4.31 (2.68)	6.56 (5.75)	3.56 (3.06)
**Digit Coding	86.35 (13.22)	58.77 (14.46)	83.29 (13.73)	62.61 (16.80)

Note. Standard deviations are presented in parentheses.

[§]Age difference between older adults in Description and Experience conditions, $p < .05$

*Older > Younger, $p < .05$

**Younger > Older, $p < .05$

Table 4

Background Measures and Cognitive Tests

Name	Measure	Rationale	Result	References
MMSE	Dementia	Screen participants	No differences	Folstein, Folstein, & McHugh (1975)
Numeracy Scale	General Numeracy	Form low and high numeracy groups	Main effect of format (regardless of numeracy group or age)	Lipkus, Samsa, & Rimer (2001)
* Digit Symbol Coding	Perceptual and Motor Speed	Quantify whether typical differences are present	YA outperformed OA (typical result)	Wechsler Intelligence Scale (subsection; 1997)
DOSPRT	Risk Taking	Judging risks may reflect probability judgments (may provide insight into estimates made)	Preliminary analysis showed no differences in age, group or format (dropped)	Weber, Blais, and Betz (2002)
* PANAS	Positive and Negative Mood	Quantify whether typical differences are present	OA more positive; YA more negative (typical result)	Watson, Clark, & Tellegen (1988)
* DASS	Depression, Anxiety and Stress	Quantify whether typical differences are present	YA reported higher depression, anxiety and stress scores than OA (typical result)	Lovibond (1995)
Self-Assessment	Self-Reflection of Performance	Analyze whether participants can accurately judge their performance	Main effect of format for Q's 2 and 3	Armstrong & Spaniol

^a One coin toss question added to the 11-item Lipkus Numeracy scale developed by Schwartz et. al. (1997).

^b Questionnaire included both Likert scale items and 3 open-ended questions.

* literature shows consistent age differences.

Table 5

Mean Absolute Judgment Error and Standard Deviations of Prevalence Estimates

	<u>Description</u>		<u>Experience</u>	
	<u>Younger</u>	<u>Older</u>	<u>Younger</u>	<u>Older</u>
Hilioplasmosis	1.87(10.78)	2.72(12.06)	5.34(14.23)	2.02(4.14)
Mytoplasia	0(0)	0(0)	7.19(16.54)	4.63(12.84)
Polykronisia	0(0)	.07(.48)	4.06(6.15)	1.48(2.91)
Zymbosis	0(0)	3.79(17.50)	4.56(13.66)	5.68(16.06)

Table 6

Mean Absolute Judgment Error and Standard Deviations of Sensitivity Estimates

	<u>Description</u>		<u>Experience</u>	
	<u>Younger</u>	<u>Older</u>	<u>Younger</u>	<u>Older</u>
Hilioplasmosis	5.15 (16.71)	1.79(8.09)	29.62(25.28)	32.76(26.53)
Mytoplasia	3.02(13.56)	7.33(21.97)	28.94(24.96)	28.87(24.77)
Polykronisia	0(0)	4.51(19.75)	36.40(42.43)	32.82(43.34)
Zymbosis	.27(1.75)	13.74(33.10)	35.62(44.63)	35.94(44.76)

Table 7

Mean Absolute Judgment Error and Standard Deviations of Specificity Estimates

	<u>Description</u>		<u>Experience</u>	
	<u>Younger</u>	<u>Older</u>	<u>Younger</u>	<u>Older</u>
Hilioplasmosis	13.72(29.35)	20.63(34.86)	9.78(14.07)	6.90(15.62)
Mytoplasia	11.65(24.72)	25.54(34.82)	6.89(14.31)	10.35(22.13)
Polykronisia	4.72(18.54)	17.34(31.92)	3.02(4.53)	5.39(13.93)
Zymbosis	10.49(25.04)	13.85(29.36)	3.88(5.74)	4.66(11.71)

Table 8

Mean Absolute Judgment Error and Standard Deviations of False-Alarm Estimates

	<u>Description</u>		<u>Experience</u>	
	<u>Younger</u>	<u>Older</u>	<u>Younger</u>	<u>Older</u>
Hilioplasmosis	2.46(14.11)	5.84(18.92)	6.49(11.07)	5.96(18.75)
Mytoplasia	4.01(15.49)	7.15(20.86)	9.28(18.29)	3.61(3.10)
Polykronisia	0(0)	9.53(27.54)	10.29(21.71)	8.44(22.12)
Zymbosis	.84(2.87)	11.33(27.74)	6.85(16.82)	5.80(13.45)

Table 9

Mean Absolute Judgment Error and Standard Deviations of Positive Predictive Value Estimates

	<u>Description</u>		<u>Experience</u>	
	<u>Younger</u>	<u>Older</u>	<u>Younger</u>	<u>Older</u>
Hilioplasmosis	31.25(13.52)	26.46(13.01)	18.52(15.58)	28.17(18.76)
Mytoplasia	35.65(21.61)	41.44(18.69)	22.14(20.88)	23.69(20.27)
Polykronisia	55.10(30.04)	55.74(30.23)	15.97(20.08)	10.98(13.51)
Zymbosis	60.65(39.38)	65.63(35.38)	8.29(16.84)	10.89(20.66)

Table 10

Mean Absolute Judgment Error and Standard Deviations of Negative Predictive Value Estimates

	<u>Description</u>		<u>Experience</u>	
	<u>Younger</u>	<u>Older</u>	<u>Younger</u>	<u>Older</u>
Hilioplasmosis	22.34(35.18)	34.27(37.11)	22.27(32.26)	9.08(18.07)
Mytoplasia	26.17(29.23)	32.89(35.95)	12.31(20.18)	11.53(23.88)
Polykronisia	21.17(36.16)	22.88(36.26)	12.36(22.28)	9.56(21.08)
Zymbosis	14.52(28.76)	26.24(37.81)	17.81(27.94)	4.98(6.48)

Figures

View series of 100 representative patients

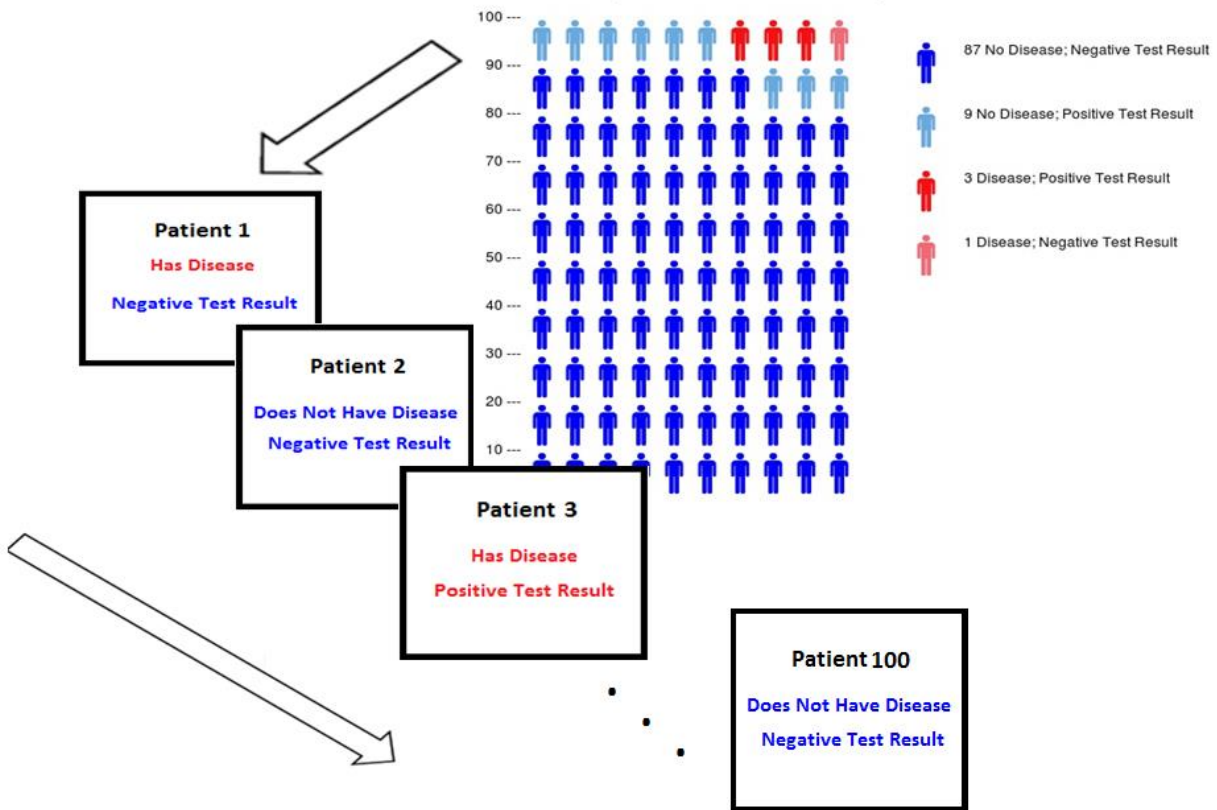


Figure 1. Schematic of experience condition. This figure illustrates the representative population participants experience through sequential presentation on the computer screen.

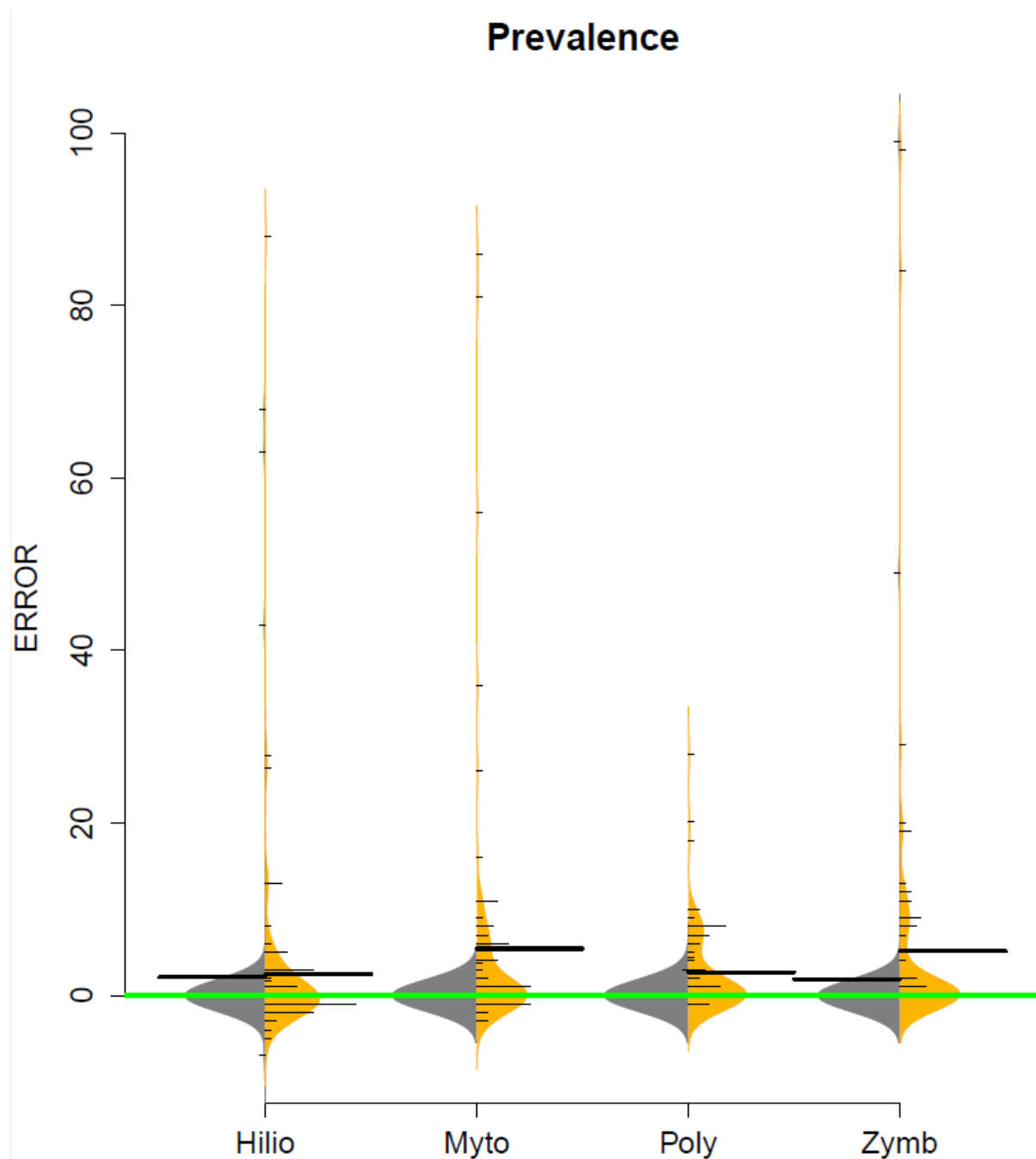


Figure 2. Estimation accuracy of prevalence. The bean plot shows estimation accuracy of prevalence. Grey areas represent participants in the description condition and gold represents participants in the experience condition. Each line plotted horizontally indicates the frequency of absolute errors made. The green line at 0 indicates accurate performance. The thick black lines indicate the mean judgment error for description and experience conditions. On the left and right sides of the plot are density functions for the distribution of answers participants provided in each condition. Plotted along the x-axis are the diseases: Hilioplasmosis, Mytoplasia, Polykronisia and Zymbosis.

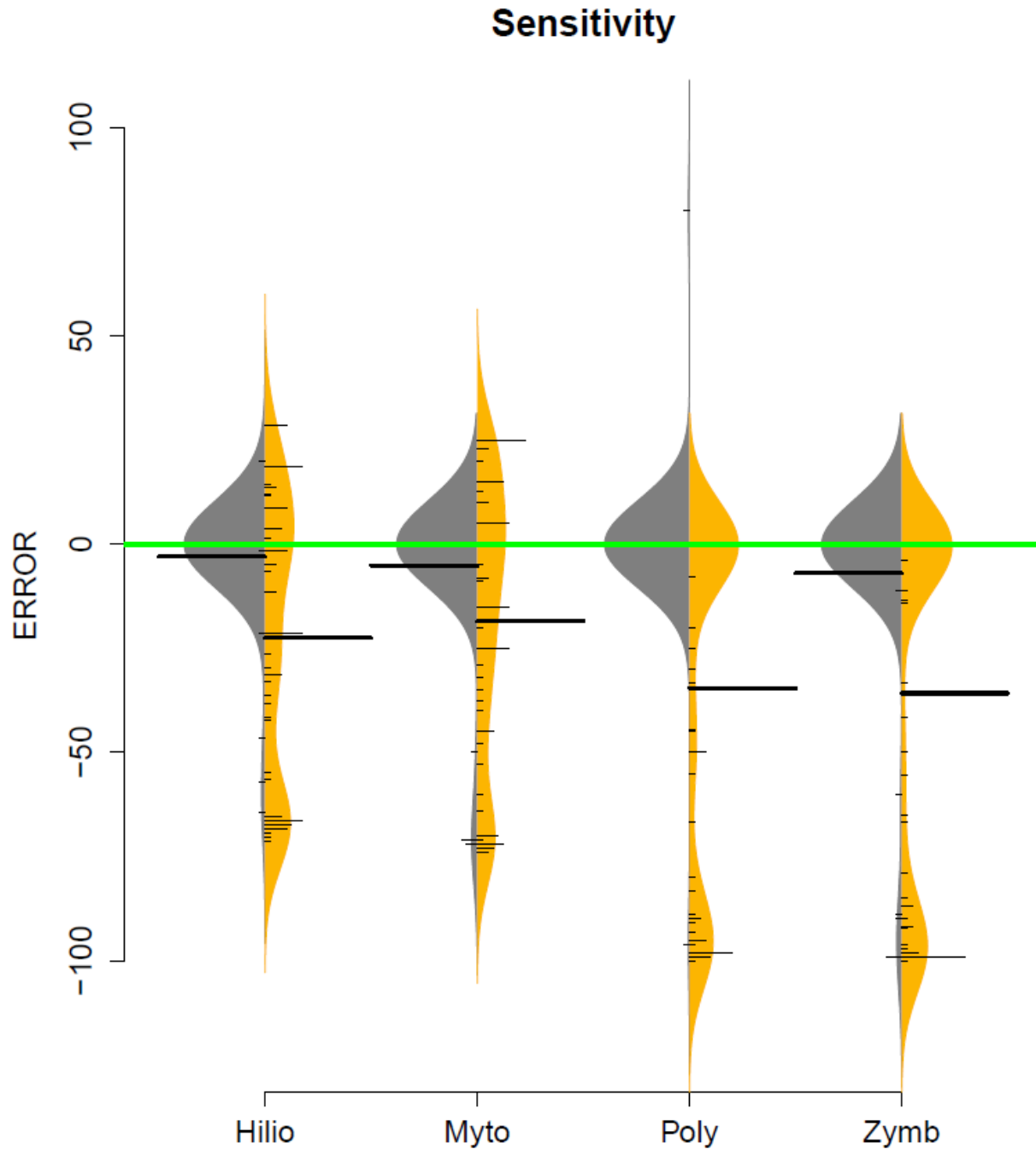


Figure 3. Estimation accuracy of sensitivity. The bean plot shows estimation accuracy of sensitivity. Grey areas represent participants in the description condition and gold represents participants in the experience condition. Each line plotted horizontally indicates the frequency of absolute errors made. The thick green line emphasizes where 0 is on the plot indicating no judgment error. The thick black lines indicate the mean judgment error in each format type, that is description and experience conditions. On the left and right sides of the plot are density functions for the distribution of answers participants provided in each condition. Plotted along the x-axis are the diseases: Hilioplasmosis, Mytoplasia, Polykronisia and Zymbosis.

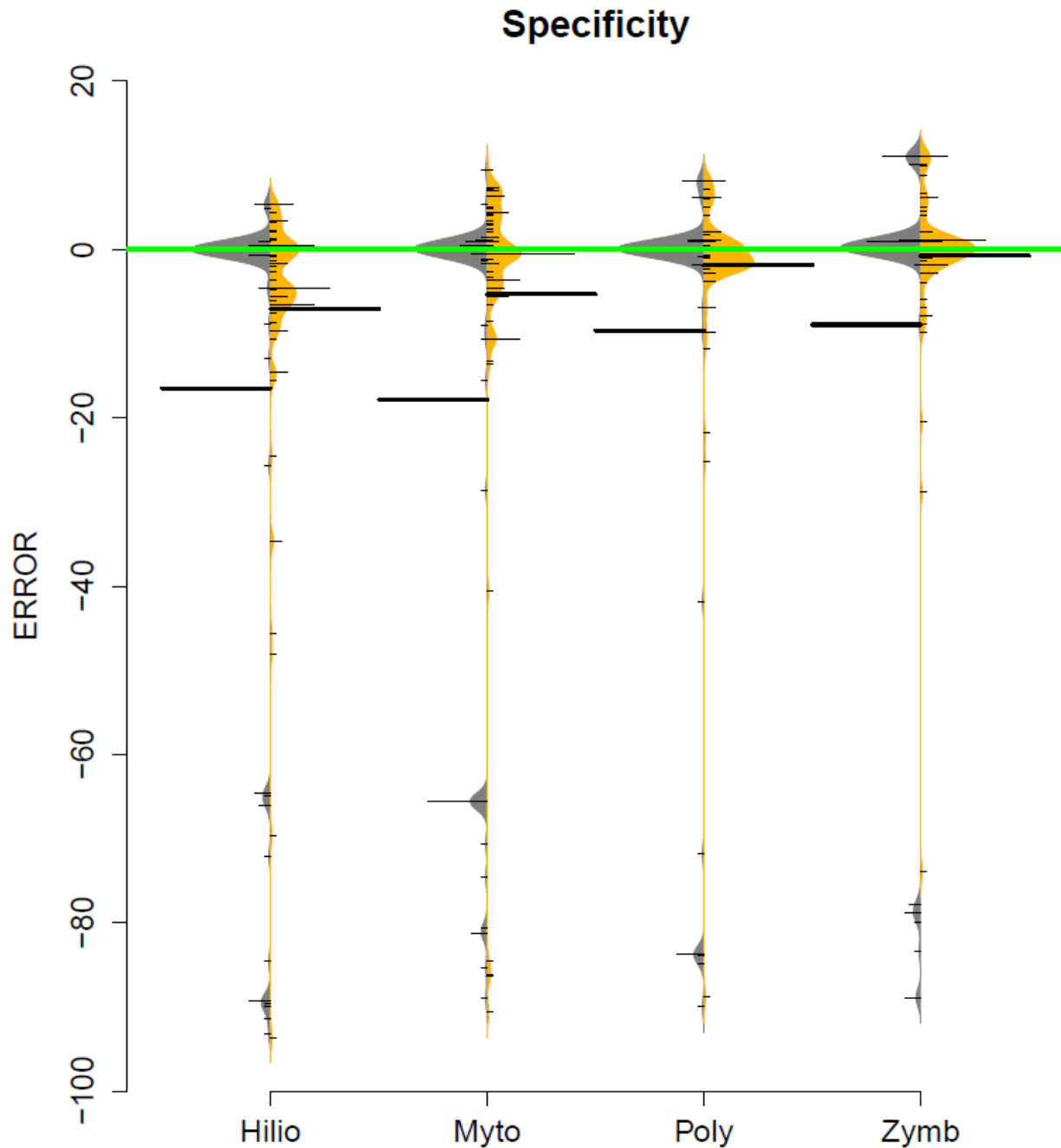


Figure 4. Estimation accuracy of specificity. The bean plot shows estimation accuracy of specificity. Grey areas represent participants in the description condition and gold represents participants in the experience condition. Each line plotted horizontally indicates the frequency of absolute errors made. The thick green line emphasizes where 0 is on the plot indicating no judgment error. The thick black lines indicate the mean judgment error in each format type, that is description and experience conditions. On the left and right sides of the plot are density functions for the distribution of answers participants provided in each condition. Plotted along the x-axis are the diseases: Hilioplasmosis, Mytoplasia, Polykronisia and Zymbosis

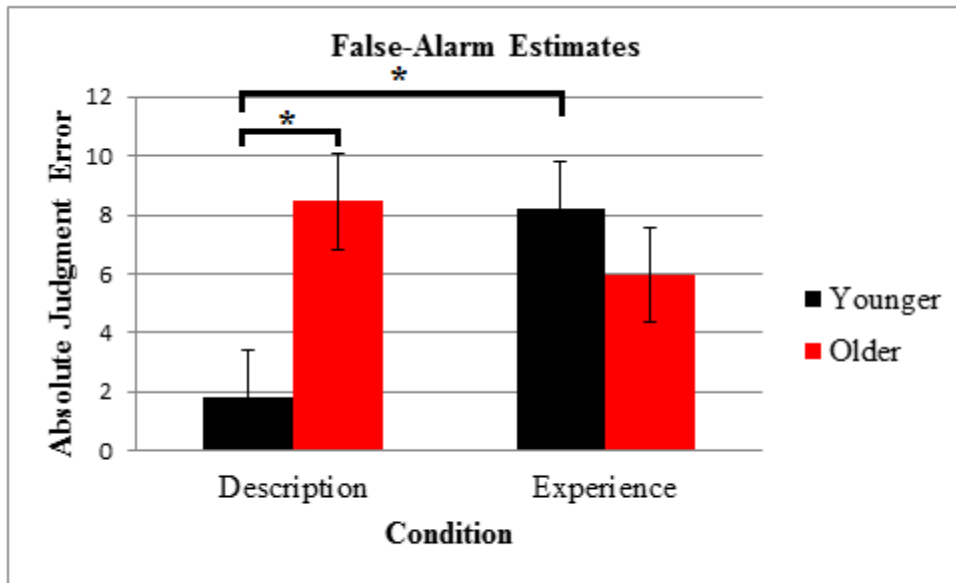


Figure 5. Bar graph illustrating age by format interaction for false-alarm estimates. The bar graph displays mean absolute judgment error and standard error bars illustrating the age by format interaction for false-alarm estimates. Younger adults made significantly more accurate false-alarm estimates in the description condition than in the experience condition. Younger adults also made significantly more accurate false-alarm estimates in the description condition compared to older adults in the description condition.

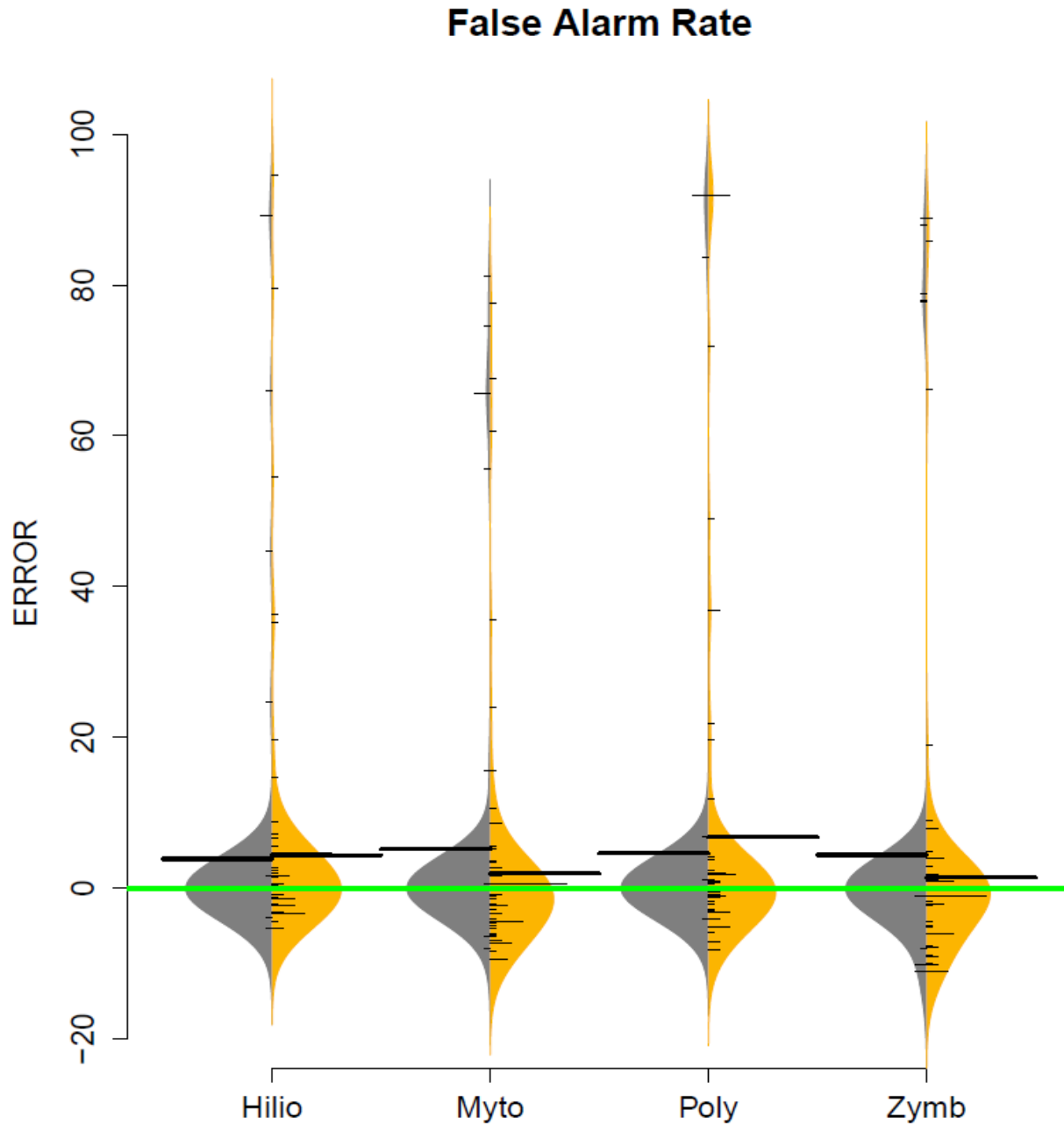


Figure 6. Estimation accuracy of false-alarm rate. The bean plot shows estimation accuracy of false-alarm rate. Grey areas represent participants in the description condition and gold represents participants in the experience condition. Each line plotted horizontally indicates the frequency of absolute errors made. The thick green line emphasizes where 0 is on the plot indicating no judgment error. The thick black lines indicate the mean judgment error in each format type that is description and experience conditions. On the left and right sides of the plot are density functions for the distribution of answers participants provided in each condition. Plotted along the x-axis are the diseases: Hilioplasmosis, Mytoplasia, Polykronisia and Zymbosis.

Positive Predictive Value

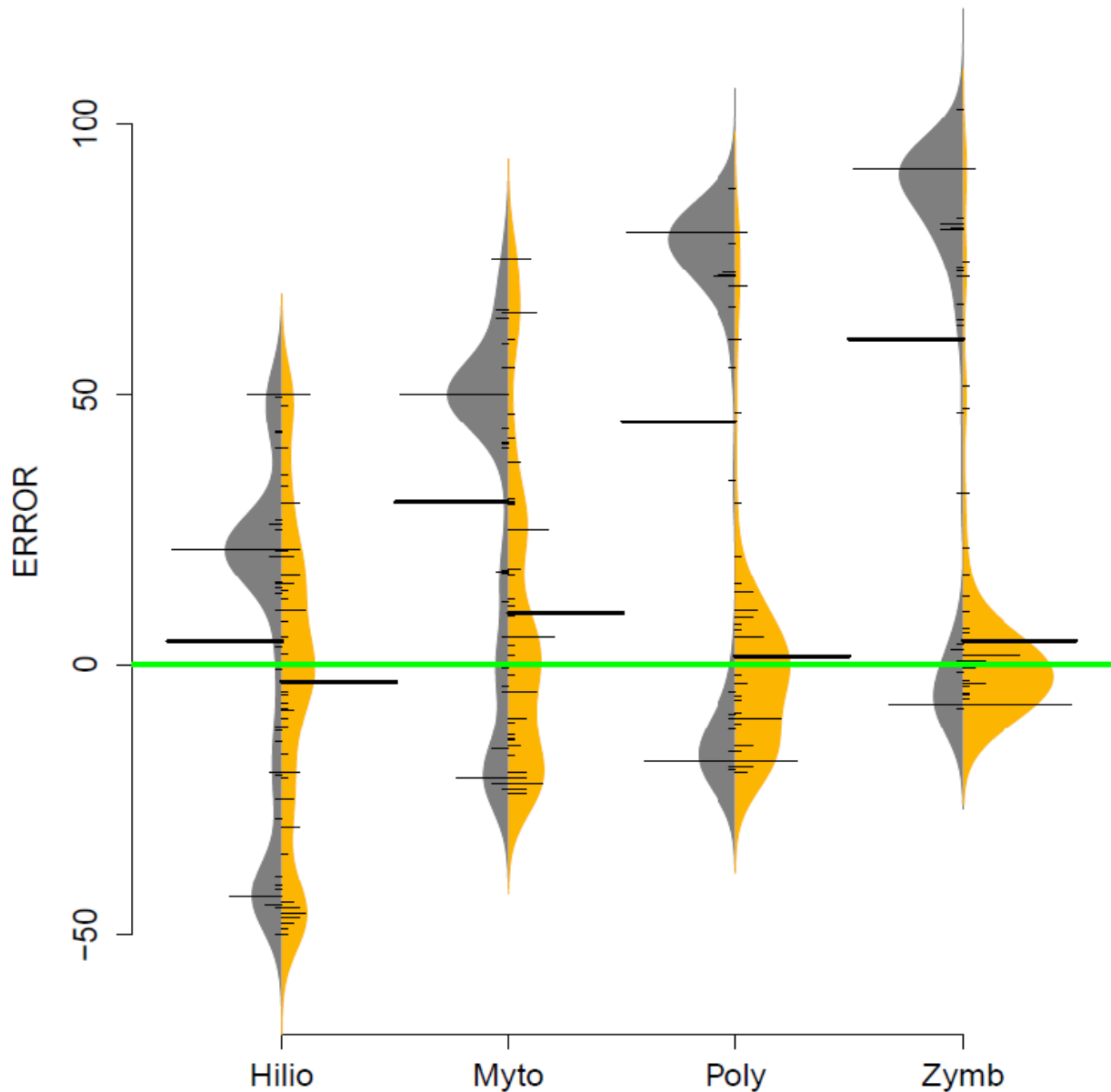


Figure 7. Estimation accuracy of PPV. The bean plot shows estimation accuracy of PPV. Grey areas represent participants in the description condition and gold represents participants in the experience condition. Each line plotted horizontally indicates the frequency of absolute errors made. The thick green line emphasizes where 0 is on the plot indicating no judgment error. The thick black lines indicate the mean judgment error in each format type, that is description and experience conditions. On the left and right sides of the plot are density functions for the distribution of answers participants provided in each condition. Plotted along the x-axis are the diseases: Hilioplasmosis, Mytoplasia, Polykronisia and Zymbosis.

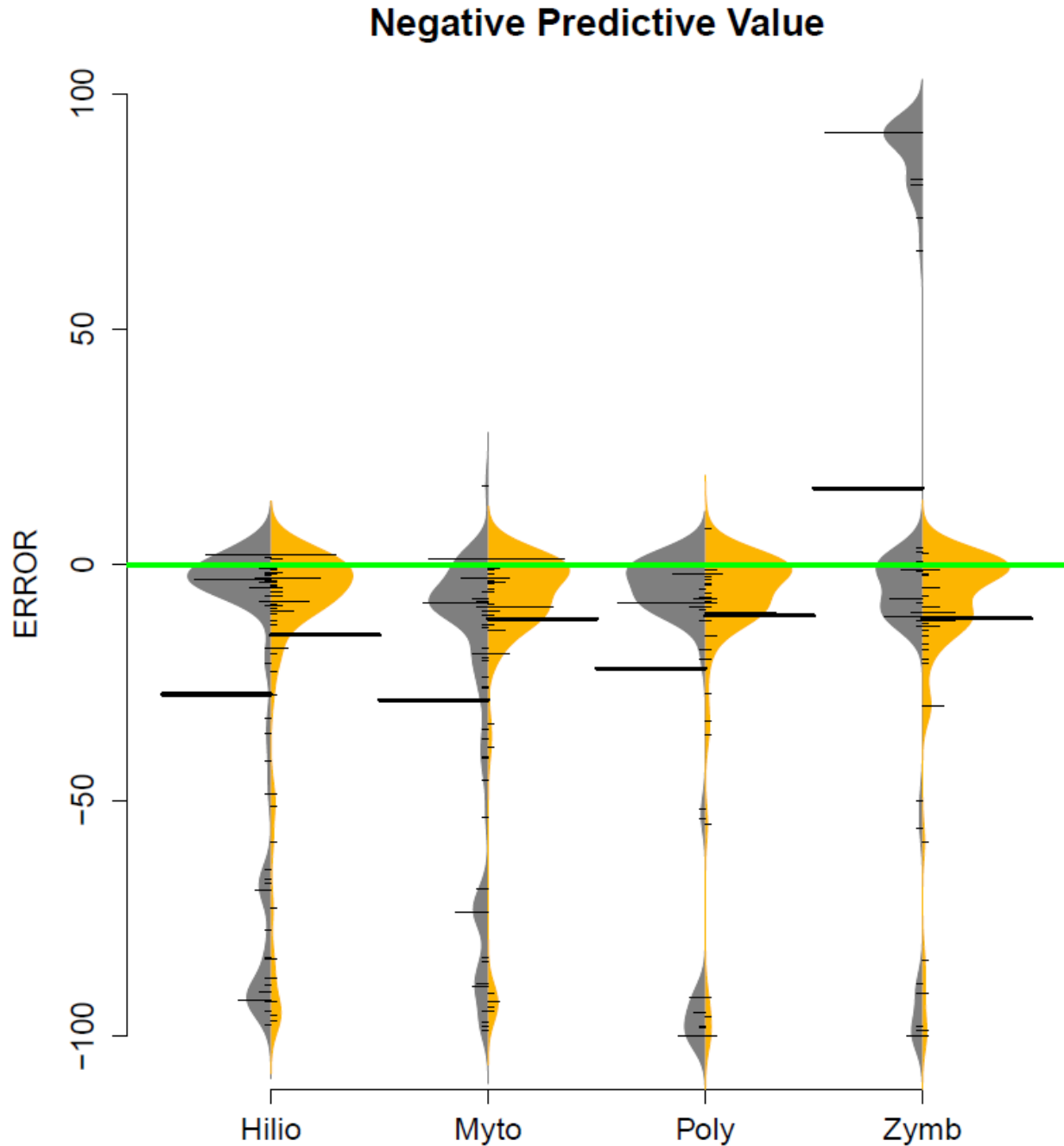


Figure 8. Estimation accuracy of NPV. The bean plot shows estimation accuracy of NPV. Grey areas represent participants in the description condition and gold represents participants in the experience condition. Each line plotted horizontally indicates the frequency of absolute errors made. The thick green line emphasizes where 0 is on the plot indicating no judgment error. The thick black lines indicate the mean judgment error in each format type, that is description and experienced conditions. On the left and right sides of the plot are density functions for the distribution of answers participants provided in each condition. Plotted along the x-axis are the diseases: Hilioplasmosis, Mytoplasia, Polykronisia and Zymbosis.

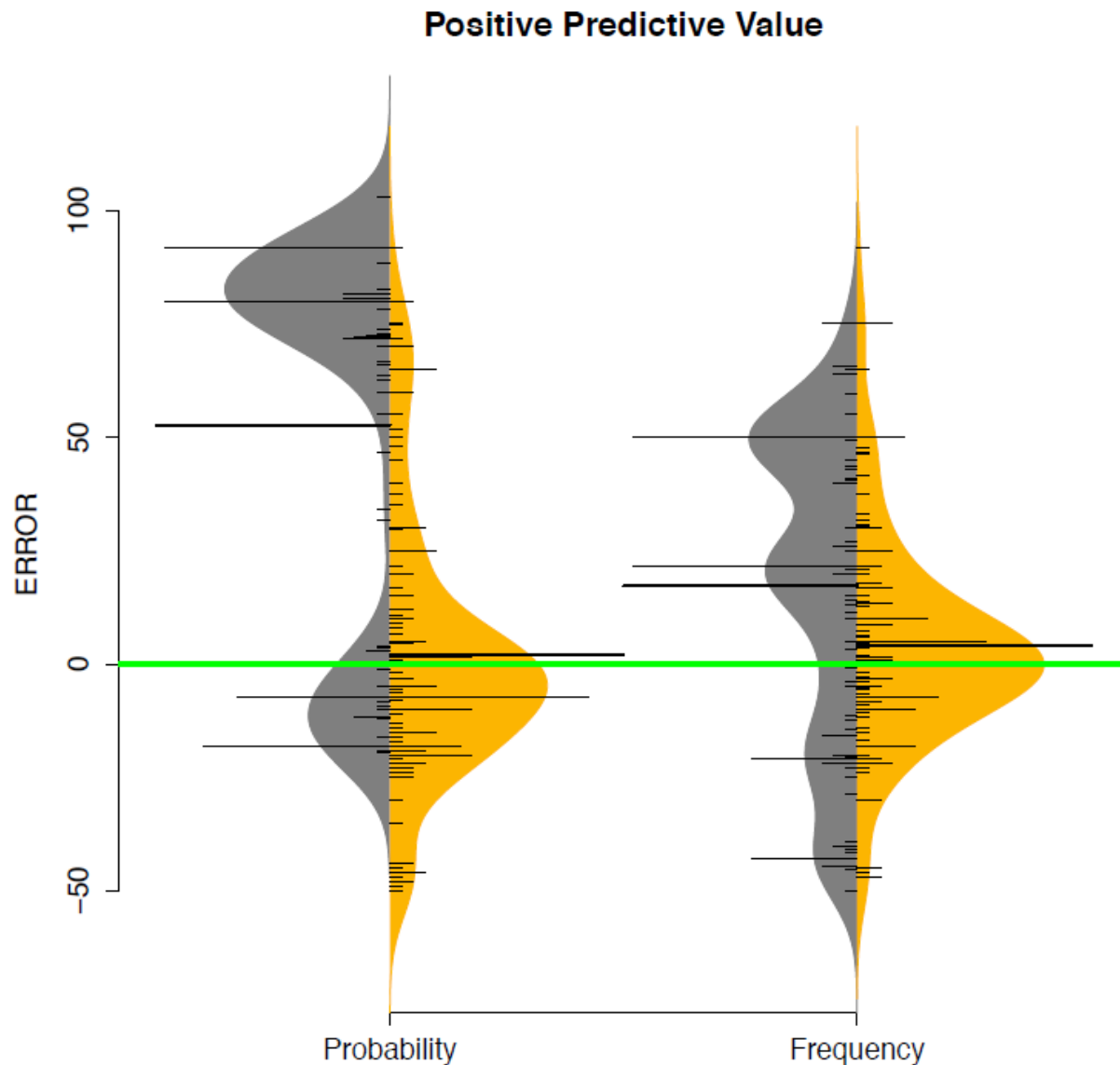


Figure 9. Accuracy of PPV estimates as a function of response type. This figure illustrates the accuracy of participants' PPV estimates when information is presented as frequencies compared to probabilities across diseases and conditions. Grey areas represent participants in the description condition and gold represents participants in the experience condition. Each line plotted horizontally indicates the frequency of absolute errors made. The thick green line emphasizes where 0 is on the plot indicating no judgment error. The thick black lines indicate the mean judgment error in each format type, description and experienced conditions respectively. On the left and right sides of the plot are density functions for the distribution of answers participants provided in each condition. Plotted along the x-axis are the response types: Probability and Frequency.

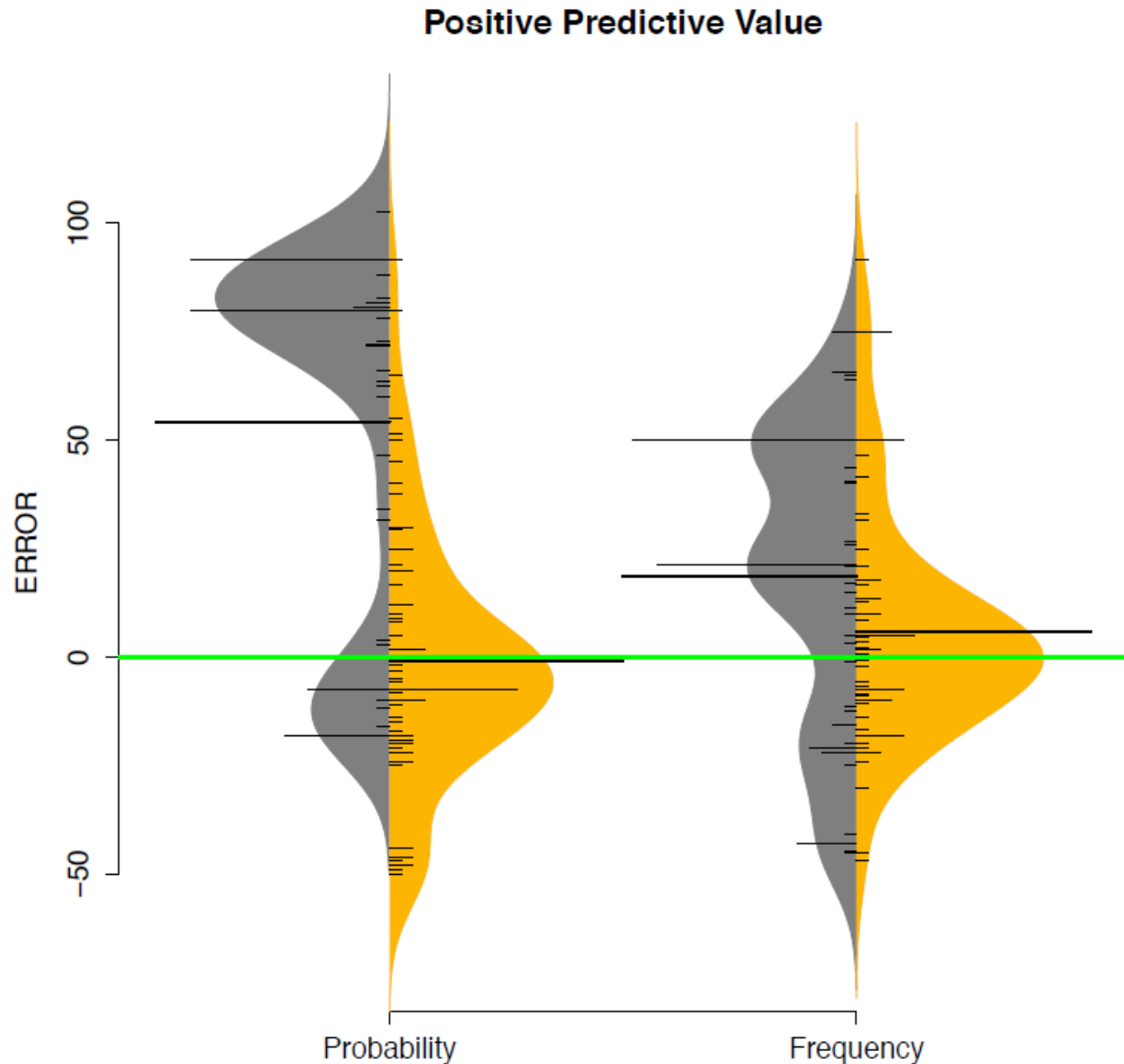


Figure 10. Accuracy of PPV estimates as a function of older adults. This figure illustrates the accuracy of older adults' PPV estimates when information is presented as frequencies compared to probabilities across diseases and conditions. Grey areas represent participants in the description condition and gold represents participants in the experience condition. Each line plotted horizontally indicates the frequency of absolute errors made. The thick green line emphasizes where 0 is on the plot indicating no judgment error. The thick black lines indicate the mean judgment error in each format type, description and experienced conditions respectively. On the left and right sides of the plot are density functions for the distribution of answers participants provided in each condition. Plotted along the x-axis are the response types: Probability and Frequency.

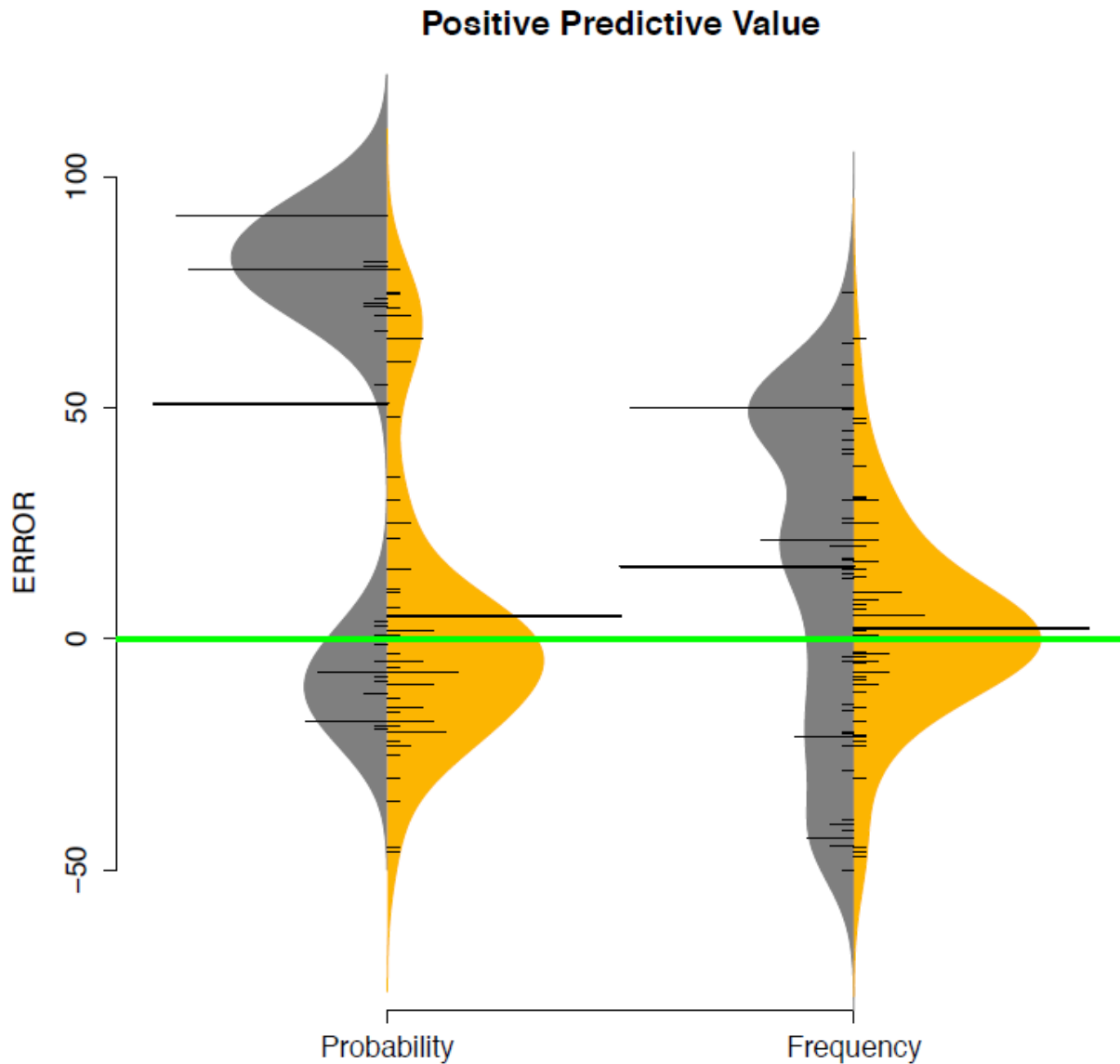


Figure 11. Accuracy of PPV estimates between response types of younger adults. This figure illustrates the accuracy of younger adults' PPV estimates when information is presented as frequencies compared to probabilities across diseases and conditions. Grey areas represent participants in the description condition and gold represents participants in the experience condition. Each line plotted horizontally indicates the frequency of absolute errors made. The thick green line emphasizes where 0 is on the plot indicating no judgment error. The thick black lines indicate the mean judgment error in each format type, that is description and experienced conditions respectively. On the left and right sides of the plot are density functions for the distribution of answers participants provided in each condition. Plotted along the x-axis are the response types: Probability and Frequency.

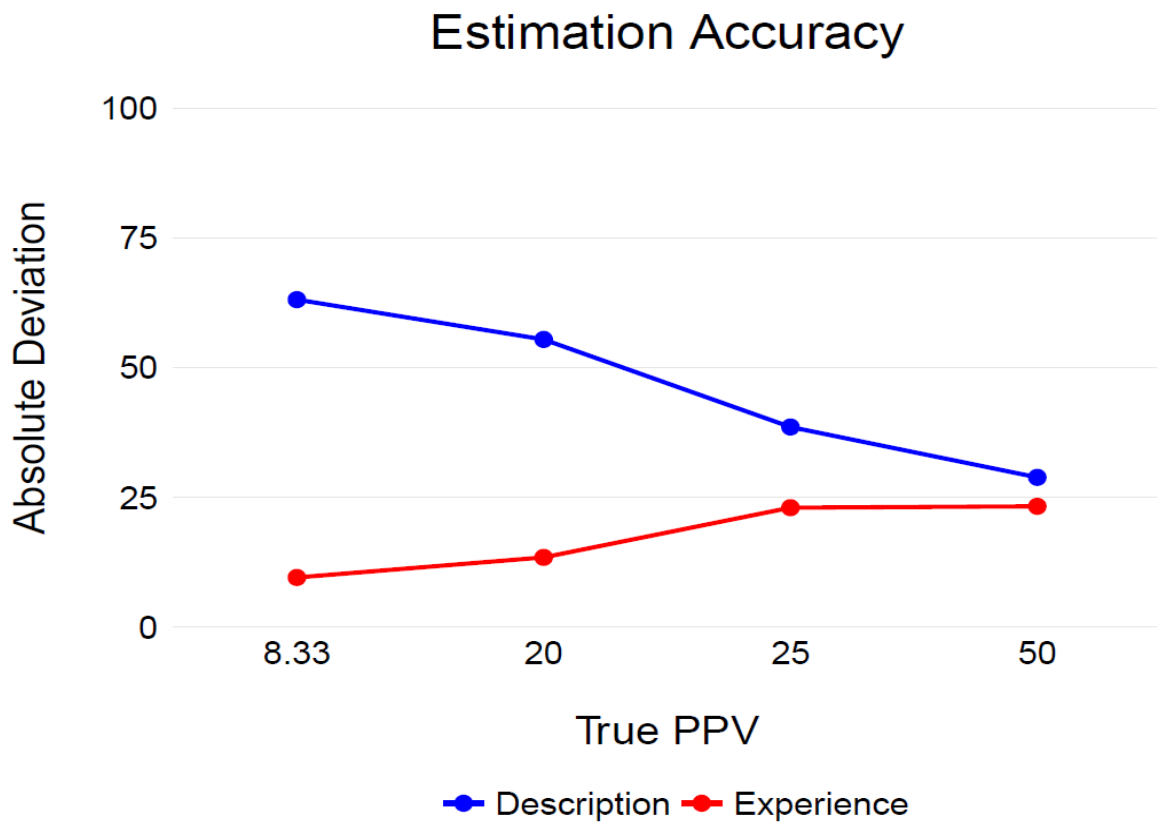


Figure 12. Accuracy of PPV estimates in experience and description condition. This figure illustrates the differences in estimate accuracy of the PPV in each condition across age groups and diseases.

Appendices

APPENDIX A: DESCRIPTION-BASED MEDICAL SCREENING TEST: HILIOPLASMOSIS

Medical Screening Test Disease: Hilioplasmosis

To determine whether a person is at risk of Hilioplasmosis, doctors sometimes conduct blood tests.

The passage below displays how common the disease is and how accurate the diagnostic tests are.

7 out of every 100 people have Hilioplasmosis.

If a person has Hilioplasmosis, it is not sure that he or she will have a positive result from the blood test. More precisely, only 71.43 of every 100 such people will have a positive result from the blood sample.

If a person does not have Hilioplasmosis, it is still possible that he or she will have a positive result from the blood sample. More precisely, 5.38 out of every 100 such people will have a positive result on the blood test even though they do not have the disease.

Please write the name of the disease and answer the questions in the booklet provided by using the information in the passage above.

Press the space bar when you have finished answering the questions.

APPENDIX B: DESCRIPTION-BASED MEDICAL SCREENING TEST: MYTOPLASIA

Medical Screening Test Disease: Mytoplasia

To determine whether a person is at risk of Mytoplasia, doctors sometimes conduct genetic testing.

The passage below displays how common the disease is and how accurate the diagnostic tests are.

4 out of every 100 people have Mytoplasia.

If a person has Mytoplasia, it is not sure that he or she will have a positive result on the genetic test. More precisely, only 75 of every 100 such people will have a positive result on the genetic test.

If a person does not have Mytoplasia, it is still possible that he or she will have a positive result on the genetic test. More precisely, 9.38 out of every 100 such people will have a positive result on the genetic test.

Please write the name of the disease and answer the questions in the booklet provided by using the information in the passage above.

Press the space bar when you have finished answering the questions.

APPENDIX C: DESCRIPTION-BASED MEDICAL SCREENING TEST: POLYKRONISIA

Medical Screening Test Disease: Polykronisia

To determine whether a person is at risk of Polykronisia, doctors sometimes conduct blood tests.

The passage below displays how common the disease is and how accurate the diagnostic tests are.

The probability that a person has Polykronisia is 2%.

If a person has Polykronisia, he or she has a 100% probability of having a positive result from the blood sample.

If a person does not have Polykronisia, it is still possible that he or she will have a positive result from the blood sample. More precisely, he or she has an 8.16% probability of having a positive result from the blood sample.

Please write the name of the disease and answer the questions in the booklet provided by using the information in the passage above.

Press the space bar when you have finished answering the questions.

APPENDIX D: DESCRIPTION-BASED MEDICAL SCREENING TEST: ZYMBOSIS

Medical Screening Test Disease: Zymbosis

To determine whether a person is at risk of Zymbosis, doctors sometimes conduct genetic testing.

The passage below displays how common the disease is and how accurate the diagnostic tests are.

The probability that a person has Zymbosis is 1%.

If a person has Zymbosis, he or she has a 100% probability of having a positive result on a genetic test.

If a person does not have Zymbosis, it is still possible that he or she will have a positive result on a genetic test. More precisely, he or she has an 11.11% probability of having a positive result on a genetic test.

Please write the name of the disease and answer the questions in the booklet provided by using the information in the passage above.

Press the space bar when you have finished answering the questions.

APPENDIX E: VERSION 1: ESTIMATION BOOKLET

Version 1

Example

Medical Screening Test

Disease Name:_____.

How many patients have the disease?

_____ %

Among patients who have the disease, how many receive a positive test result?

_____ %

Among patients who do not have the disease, how many receive a negative test result?

_____ %

Among patients who do not have the disease, how many receive a positive test result?

_____ %

Among patients who received a positive test result, how many have the disease?

_____ %

Among patients who received a negative test result, how many do not have the disease?

_____ %

APPENDIX E: VERSION 1: ESTIMATION BOOKLET

Version 1

Medical Screening Test 1

Disease Name:_____.

How many patients have the disease?

_____ out of _____

Among patients who have the disease, how many receive a positive test result?

_____ out of _____

Among patients who do not have the disease, how many receive a negative test result?

_____ out of _____

Among patients who do not have the disease, how many receive a positive test result?

_____ out of _____

Among patients who received a positive test result, how many have the disease?

_____ out of _____

Among patients who received a negative test result, how many do not have the disease?

_____ out of _____

APPENDIX E: VERSION 1: ESTIMATION BOOKLET

Version 1

Medical Screening Test 2

Disease Name:_____.

How many patients have the disease?

_____ %

Among patients who have the disease, how many receive a positive test result?

_____ %

Among patients who do not have the disease, how many receive a negative test result?

_____ %

Among patients who do not have the disease, how many receive a positive test result?

_____ %

Among patients who received a positive test result, how many have the disease?

_____ %

Among patients who received a negative test result, how many do not have the disease?

_____ %

APPENDIX E: VERSION 1: ESTIMATION BOOKLET

Version 1

Medical Screening Test 3

Disease Name:_____.

How many patients have the disease?

_____ out of _____

Among patients who have the disease, how many receive a positive test result?

_____ out of _____

Among patients who do not have the disease, how many receive a negative test result?

_____ out of _____

Among patients who do not have the disease, how many receive a positive test result?

_____ out of _____

Among patients who received a positive test result, how many have the disease?

_____ out of _____

Among patients who received a negative test result, how many do not have the disease?

_____ out of _____

APPENDIX E: VERSION 1: ESTIMATION BOOKLET

Version 1

Medical Screening Test 4

Disease Name:_____.

How many patients have the disease?

_____ %

Among patients who have the disease, how many receive a positive test result?

_____ %

Among patients who do not have the disease, how many receive a negative test result?

_____ %

Among patients who do not have the disease, how many receive a positive test result?

_____ %

Among patients who received a positive test result, how many have the disease?

_____ %

Among patients who received a negative test result, how many do not have the disease?

_____ %

APPENDIX F: VERSION 2: ESTIMATION BOOKLET

Version 2

Example

Medical Screening Test

Disease Name:_____.

How many patients have the disease?

_____ out of _____

Among patients who have the disease, how many receive a positive test result?

_____ out of _____

Among patients who do not have the disease, how many receive a negative test result?

_____ out of _____

Among patients who do not have the disease, how many receive a positive test result?

_____ out of _____

Among patients who received a positive test result, how many have the disease?

_____ out of _____

Among patients who received a negative test result, how many do not have the disease?

_____ out of _____

APPENDIX F: VERSION 2: ESTIMATION BOOKLET

Version 2

Medical Screening Test 1

Disease Name:_____.

How many patients have the disease?

_____ %

Among patients who have the disease, how many receive a positive test result?

_____ %

Among patients who do not have the disease, how many receive a negative test result?

_____ %

Among patients who do not have the disease, how many receive a positive test result?

_____ %

Among patients who received a positive test result, how many have the disease?

_____ %

Among patients who received a negative test result, how many do not have the disease?

_____ %

APPENDIX F: VERSION 2: ESTIMATION BOOKLET

Version 2

Medical Screening Test 2

Disease Name:_____.

How many patients have the disease?

_____ out of _____

Among patients who have the disease, how many receive a positive test result?

_____ out of _____

Among patients who do not have the disease, how many receive a negative test result?

_____ out of _____

Among patients who do not have the disease, how many receive a positive test result?

_____ out of _____

Among patients who received a positive test result, how many have the disease?

_____ out of _____

Among patients who received a negative test result, how many do not have the disease?

_____ out of _____

APPENDIX F: VERSION 2: ESTIMATION BOOKLET

Version 2

Medical Screening Test 3

Disease Name:_____.

How many patients have the disease?

_____ %

Among patients who have the disease, how many receive a positive test result?

_____ %

Among patients who do not have the disease, how many receive a negative test result?

_____ %

Among patients who do not have the disease, how many receive a positive test result?

_____ %

Among patients who received a positive test result, how many have the disease?

_____ %

Among patients who received a negative test result, how many do not have the disease?

_____ %

APPENDIX F: VERSION 2: ESTIMATION BOOKLET

Version 2

Medical Screening Test 4

Disease Name:_____.

How many patients have the disease?

_____ out of _____

Among patients who have the disease, how many receive a positive test result?

_____ out of _____

Among patients who do not have the disease, how many receive a negative test result?

_____ out of _____

Among patients who do not have the disease, how many receive a positive test result?

_____ out of _____

Among patients who received a positive test result, how many have the disease?

_____ out of _____

Among patients who received a negative test result, how many do not have the disease?

_____ out of _____

APPENDIX G: MINI-MENTAL STATE EXAMINATION

FOLSTEIN MINI MENTAL STATE EXAM

ORIENTATION

- What year is it? _____(1) What season is it? _____(1)
- What is today's date? _____(2)
(prompt for month and/or date if either omitted)
- What day of the week is it? _____ (1)
- What province are we in? _____(1)
- What country are we in? _____(1)
- What city are we in? _____ (1)
- What hospital/bldg. is this? _____ (1)
- What floor are we on? _____ (1)

REGISTRATION

- Now I am going to say three words. When I stop, I want you to say them after me. “Bus” “Door” “Rose” (Score one point for each word correct on the first trial. Repeat until all words are repeated correctly, and record the number of trials.) (3)

ATTENTION AND CALCULATION

- I am going to say a word. I want you to spell the word backwards. The word is 'WORLD'. (5)

RECALL

- What were those three words I had you repeat after me? (record) (3)

LANGUAGE

- What is this? (hold up a pencil) _____ And what is this?
(point to a watch) _____ (2)
- Repeat this phrase after me: No ifs, ands, or buts. (1)
- I am going to hand you a piece of paper. I want you to
take it in your right hand, fold it in half, and lay it on

SCORE

your lap. (hand paper only after giving all commands) (3)

- Read this aloud and do what it says. (1)
- I want you to write a complete sentence. (1)
- Copy this design as carefully as you can. (1)

APPENDIX H: NUMERACY QUESTIONNAIRE

Numeracy Questionnaire

Please answer the following questions to the best of your ability.

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

___ 1 in 100 ___ 1 in 1000 **X** 1 in 10

2. Which of the following represents the biggest risk of getting a disease?

___ 1% ___ 10% **X** 15%

3. If the chance of getting a disease were 10%, how many people would be expected to get the disease out of 100? **10**
4. If the chance of getting a disease were 10%, how many people would be expected to get the disease out of 1000? **100**
5. If the chance of getting a disease were 20 out of 100, this would be the same as having a **20%** chance of getting the disease.
6. If Person A's risk of getting a disease is 1% in ten years, and Person B's risk is double that of A's, what is B's risk? **2%**
7. If Person A's chance of getting a disease is 1 in 100 in ten years, and Person B's risk is double that of A, what is B's risk? **2 out of 100** (*I also accept it if they reduced the fraction, i.e., 1 out of 50*)
8. In the BIG BUCKS LOTTERY, the chances of winning a \$10.00 prize are 1%. What is your best guess about how many people would win a \$10.00 prize if 1000 people each buy a single ticket from BIG BUCKS? **10 people**
9. Imagine that we roll a fair, six-sided dice 1000 times. Out of 1000 rolls, how many times do you think the die would come up even (2, 4, or 6)? **500**

10. The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected? **5 people**
11. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car? **0.1%**
12. Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips? **500** times out of 1,000.

APPENDIX I: DOMAIN SPECIFIC RISK-TAKING QUESTIONNAIRE

For each of the following statements, please indicate the **likelihood** that you would engage in the described activity or behavior if you were to find yourself in that situation. Provide a rating from *Extremely Unlikely* to *Extremely Likely*, using the following scale. Please indicate the number that best represents your answer in the space provided for that activity or behaviour:

1	2	3	4	5	6	7
Extremely Unlikely	Moderately Unlikely	Somewhat Unlikely	Not Sure	Somewhat Likely	Moderately Likely	Extremely Likely

1. Admitting that your tastes are different from those of a friend. _____
2. Going camping in the wilderness. _____
3. Betting a day's income at the horse races. _____
4. Investing 10% of your annual income in a moderate growth mutual fund. _____
5. Drinking heavily at a social function. _____
6. Taking some questionable deductions on your income tax return. _____
7. Disagreeing with an authority figure on a major issue. _____
8. Betting a day's income at a high-stake poker game. _____
9. Having an affair with a married man/woman. _____
10. Passing off somebody else's work as your own. _____
11. Going down a ski run that is beyond your ability. _____
12. Investing 5% of your annual income in a very speculative stock. _____
13. Going whitewater rafting at high water in the spring. _____
14. Betting a day's income on the outcome of a sporting event. _____
15. Engaging in unprotected sex. _____
16. Revealing a friend's secret to someone else. _____
17. Driving a car without wearing a seat belt. _____
18. Investing 10% of your annual income in a new business venture. _____
19. Taking a skydiving class. _____
20. Riding a motorcycle without a helmet. _____
21. Choosing a career that you truly enjoy over a more secure one. _____
22. Speaking your mind about an unpopular issue in a meeting at work. _____
23. Sunbathing without sunscreen. _____
24. Bungee jumping off a tall bridge. _____
25. Piloting a small plane. _____
26. Walking home alone at night in an unsafe area of town. _____
27. Moving to a city far away from your extended family. _____
28. Starting a new career in your mid-thirties. _____
29. Leaving your young children alone at home while running an errand. _____
30. Not returning a wallet you found that contains \$200. _____

People often see some risk in situations that contain uncertainty about what the outcome or consequences will be and for which there is the possibility of negative consequences. However, riskiness is a very personal and intuitive notion, and we are interested in **your gut level assessment of how risky** each situation or behavior is.

For each of the following statements, please indicate **how risky you perceive** each situation. Provide a rating from *Not at all Risky* to *Extremely Risky*, using the following scale. Please indicate the number that best represents your answer in the space provided for that activity or behaviour:

1	2	3	4	5	6	7
Not at all Risky	Slightly Risky	Somewhat Risky	Moderately Risky	Risky	Very Risky	Extremely Risky

31. Admitting that your tastes are different from those of a friend. _____
32. Going camping in the wilderness. _____
33. Betting a day's income at the horse races. _____
34. Investing 10% of your annual income in a moderate growth mutual fund. _____
35. Drinking heavily at a social function. _____
36. Taking some questionable deductions on your income tax return. _____
37. Disagreeing with an authority figure on a major issue. _____
38. Betting a day's income at a high-stake poker game. _____
39. Having an affair with a married man/woman. _____
40. Passing off somebody else's work as your own. _____
41. Going down a ski run that is beyond your ability. _____
42. Investing 5% of your annual income in a very speculative stock. _____
43. Going whitewater rafting at high water in the spring. _____
44. Betting a day's income on the outcome of a sporting event. _____
45. Engaging in unprotected sex. _____
46. Revealing a friend's secret to someone else. _____
47. Driving a car without wearing a seat belt. _____
48. Investing 10% of your annual income in a new business venture. _____
49. Taking a skydiving class. _____
50. Riding a motorcycle without a helmet. _____
51. Choosing a career that you truly enjoy over a more secure one. _____
52. Speaking your mind about an unpopular issue in a meeting at work. _____
53. Sunbathing without sunscreen. _____
54. Bungee jumping off a tall bridge. _____
55. Piloting a small plane. _____
56. Walking home alone at night in an unsafe area of town. _____
57. Moving to a city far away from your extended family. _____
58. Starting a new career in your mid-thirties. _____
59. Leaving your young children alone at home while running an errand. _____
60. Not returning a wallet you found that contains \$200. _____

For each of the following statements, please indicate **the benefits** you would obtain from each situation. Provide a rating from **1 to 7**, using the following scale. Please indicate the number that best represents your answer in the space provided for that activity or behaviour:

1	2	3	4	5	6	7
No benefits At All			Moderate Benefits			Great Benefits

61. Admitting that your tastes are different from those of a friend. _____
62. Going camping in the wilderness. _____
63. Betting a day's income at the horse races. _____
64. Investing 10% of your annual income in a moderate growth mutual fund. _____
65. Drinking heavily at a social function. _____
66. Taking some questionable deductions on your income tax return. _____
67. Disagreeing with an authority figure on a major issue. _____
68. Betting a day's income at a high-stake poker game. _____
69. Having an affair with a married man/woman. _____
70. Passing off somebody else's work as your own. _____
71. Going down a ski run that is beyond your ability. _____
72. Investing 5% of your annual income in a very speculative stock. _____
73. Going whitewater rafting at high water in the spring. _____
74. Betting a day's income on the outcome of a sporting event. _____
75. Engaging in unprotected sex. _____
76. Revealing a friend's secret to someone else. _____
77. Driving a car without wearing a seat belt. _____
78. Investing 10% of your annual income in a new business venture. _____
79. Taking a skydiving class. _____
80. Riding a motorcycle without a helmet. _____
81. Choosing a career that you truly enjoy over a more secure one. _____
82. Speaking your mind about an unpopular issue in a meeting at work. _____
83. Sunbathing without sunscreen. _____
84. Bungee jumping off a tall bridge. _____
85. Piloting a small plane. _____
86. Walking home alone at night in an unsafe area of town. _____
87. Moving to a city far away from your extended family. _____
88. Starting a new career in your mid-thirties. _____
89. Leaving your young children alone at home while running an errand. _____
90. Not returning a wallet you found that contains \$200. _____

APPENDIX J: POSITIVE AND NEGATIVE AFFECT SCALE

PANAS-Short

This scale consists of a number of words and phrases that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate how you feel RIGHT NOW (that is, at the present moment). Use the following scale to record your answers:

1 very slightly or not at all	2 a little	3 moderately	4 quite a bit	5 extremely
	_____	afraid	_____	hostile
	_____	active	_____	excited
	_____	scared	_____	guilty
	_____	alert	_____	inspired
	_____	nervous	_____	ashamed
	_____	attentive	_____	interested
	_____	jittery	_____	upset
	_____	determined	_____	proud
	_____	irritable	_____	distressed
	_____	enthusiastic	_____	strong

APPENDIX K: DEPRESSION, ANXIETY, AND STRESS SCALE

Depression Anxiety Stress Scales

The rating scale is as follows:

0	Did not apply to me at all				
1	Applied to me to some degree, or some of the time				
2	Applied to me to a considerable degree, or a good part of time				
3	Applied to me very much, or most of the time				
1	I found it hard to wind down	0	1	2	3
2	I was aware of dryness of my mouth	0	1	2	3
3	I couldn't seem to experience any positive feeling at all	0	1	2	3
4	I experienced breathing difficulty (e.g., excessively rapid breathing, breathlessness in the absence of physical exertion)	0	1	2	3
5	I found it difficult to work up the initiative to do things	0	1	2	3
6	I tended to over-react to situations	0	1	2	3
7	I experienced trembling (e.g., in the hands)	0	1	2	3
8	I felt that I was using a lot of nervous energy	0	1	2	3
9	I was worried about situations in which I might panic and make a fool of myself	0	1	2	3
10	I felt that I had nothing to look forward to	0	1	2	3
11	I found myself getting agitated	0	1	2	3
12	I found it difficult to relax	0	1	2	3
13	I felt down-hearted and blue	0	1	2	3
14	I was intolerant of anything that kept me from getting on with what I was doing	0	1	2	3
15	I felt I was close to panic	0	1	2	3
16	I was unable to become enthusiastic about anything	0	1	2	3
17	I felt I wasn't worth much as a person	0	1	2	3
18	I felt that I was rather touchy	0	1	2	3
19	I was aware of the action of my heart in the absence of physical exertion (e.g., sense of heart rate increase, heart missing a beat)	0	1	2	3
20	I felt scared without any good reason	0	1	2	3
21	I felt that life was meaningless	0	1	2	3

Lovibond, S.H., & Lovibond, P.F., (1995). *Manual for the Depression Anxiety Stress Scales*. Sydney: The Psychology Foundation of Australia.

Lovibond, P.F., & Lovibond, S.H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33, 335-343.

APPENDIX L: DIGIT CODING TASK

VERSION 1

1	2	3	4	5	6	7	8	9
—	⊥	⊐	└	┘	○	∧	×	=

Sample Items

2	1	3	7	2	4	8	2	1	3	2	1	4	2	3	5	2	3	1	4
5	6	3	1	4	1	5	4	2	7	6	3	5	7	2	8	5	4	6	3
7	2	8	1	9	5	8	4	7	3	6	2	5	1	9	2	8	3	7	4
6	5	9	4	8	3	7	2	6	1	5	4	6	3	7	9	2	8	1	7
9	4	6	8	5	9	7	1	8	5	2	9	4	8	6	3	7	9	8	6
2	7	3	6	5	1	9	8	4	5	7	3	1	4	8	7	9	1	4	5
7	1	8	2	9	3	6	7	2	8	5	2	3	1	4	8	4	2	7	6

APPENDIX M: SELF-ASSESSMENT QUESTIONNAIRE

Estimating Properties of a Diagnostic Test

1. I am confident and comfortable working with numbers.

1	2	3	4	5
Strongly Disagree	Disagree	Do not Agree or Disagree	Agree	Strongly Agree

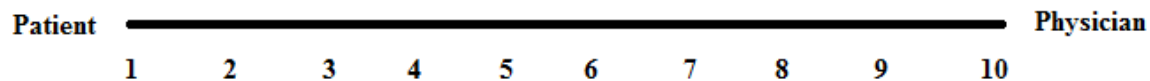
2. I found the task of estimating properties of a diagnostic test difficult.

1	2	3	4	5
Strongly Disagree	Disagree	Do not Agree or Disagree	Agree	Strongly Agree

3. I believe the estimations I made are close to the correct answers.

1	2	3	4	5
Strongly Disagree	Disagree	Do not Agree or Disagree	Agree	Strongly Agree

4. Who do you feel most comfortable with to make final decisions about your medical care – you as the patient or the physician? Please circle the number you feel best represents who you would prefer to make final medical decisions about your health care on the number line below.



5. Did you use any strategies to make the estimations? If yes, please explain.

6. How difficult did you find this experiment?

7. Please utilize the space below if you have any comments about the experiment.

References

- Alvarez, J. A. & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review*, 16, 17-42.
- Bernardo, J. M. (2003). Bayesian Statistics. *Probability and Statistics*, 21, 1 – 46.
- Brown, S. C., & Park, D. C. (2002). Roles of age and familiarity in learning of health information. *Educational Gerontology*, 28, 695-710.
- Cabeza, R. (2008). Role of parietal regions in episodic memory retrieval: the dual attentional processes hypothesis. *Neuropsychologia*, 46, 1813-1827.
- Cabeza, R., & Dennis, N.A. (2013). Frontal lobes and aging: Deterioration and compensation. In D.T. Stuss & R. T. Knight (Eds). *Principles of frontal lobe function*, 2nd Edition. Oxford University Press, New York.
- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58, 1-73.
- Daselaar, S. M., Fleck, M. S., & Cabeza, R. (2006). Triple dissociation in the medial temporal lobes: recollection, familiarity, and novelty. *Journal of Neurophysiology*, 96, 1902-1911.
- Deary, I.J., Corley, J., Gow, A. J., Harris, S.E., Houlihan, L. M., Marioni, R. E., Penke, L., Rafnsson, S. B., & Starr, J.M. (2009). Age-associated cognitive decline. *British Medical Bulletin: Oxford Journals*, 92, 135-152.
- Dew, I. T. Z. & Cabeza, R. (2011). The porous boundaries between explicit and implicit memory: behavioural and neural evidence. *Annals of the New York Academy of Sciences*, 1224, 174-190.
- Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in Cognitive Science*, 11,

379-386.

Elliott, R. (2003). Executive functions and their disorders. *British Medical Bulletin*, 65, 49–59.

Feinberg, S. E. (2006). How to report statistics in medicine (2nd Ed). *American College of Physicians*, 1 – 490.

Folstein, M., Folstein, S.E., & McHugh, P.R. (1975). “Mini-Mental State” a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189-198.

Galesic, M., Gigerenzer, G., & Straubinger, N. (2009). Natural frequencies help older adults and people with low numeracy to evaluate medical screening tests. *Medical Decision Making*, 29, 368 – 371.

Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. (2010). Do icon arrays help reduce denominator neglect? *Medical Decision Making*, 30, 672 – 684.

Garcia-Retamero, R., & Galesic, M. (2010). Who profits from visual aids: Overcoming challenges in people’s understanding of risks. *Social Science & Medicine*, 70, 1019-1025.

Gazmararian, J. A., J. W. Curran, R. M. Parker, J. M. Bernhardt, & B. A. DeBuono. (2005). Public health literacy in America: An ethical imperative. *American Journal of Preventative Medicine*, 28, 317-322.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: beyond heuristics and biases. *European Review of Social Psychology*, 2, 83-115.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2008). Helping doctors and patients make sense of health statistics. *Association Psychological Science*, 8, 53 –96.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction:

- frequency formats. *Psychological Review*, 4, 684-704.
- Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. *Trends in Cognitive Sciences*, 8, 371-377.
- Graf, P., & Ryan, L. (1990). Transfer-appropriate processing for implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 978-992.
- Halter, J. B. (1999). The challenge of communicating health information to elderly patients: A view from geriatric medicine. In D. C. Park, R. W. Morrell, & K. Shiffrin (Eds.), *Processing of medical information in aging patients: Cognitive and human factors perspectives*, 23-28. Mahwah, NJ: Erlbaum.
- Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 108, 356-358.
- Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency occurrence. *American Psychologist*, 39, 1372-1388.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *The Psychology of Learning and Motivation*, 22, 193-219.
- Hertwig, R., Barron, G., Weber, E.U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15, 534-539.
- Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, 13, 517-523.
- Hogarth, R., & Soyer, E. (2014). Providing information for decision making: Contrasting description and simulation. *Journal of Applied Research in Memory and Cognition*.
- Hogarth, R., & Soyer, E. (2010). Experiencing simulated outcomes. Barcelona Economics Working Paper Series, 1-63.

- Institute of Medicine. (2004). Health literacy: A prescription to end confusion. Washington, DC: National Academies Press.
- Irwig, L., Tosteson, A., Gatsonis, C., Lau, J., Colditz, G., Chalmers, T. C., & Mosteller, F. (1994). Guidelines for meta-analyses evaluating diagnostic tests. *Annals of Internal Medicine*, 120, 666 – 676.
- Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513-541.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110, 306-340.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology*, 3, 430-454.
- Kinoshita, S. (2001). The role of involuntary aware memory in the implicit stem and fragment completion tasks: a selective review. *Psychonomic Bulletin and Review*, 8, 58-69.
- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21, 37 – 44.
- Liu, L. L., & Gonzalez, R. (2007). Judgment and decision processes in older adults' compliance with medical regimens. *Medical Adherence and Aging: Social and Cognitive Perspectives* (1st ed). *American Psychological Association*. 201 – 232.
- Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales*. Sydney: Psychology Foundation.
- Mandler, G. (1980). Recognizing: the judgment of previous occurrence. *Psychological Review*, 87, 252-271.
- Monsell S (2003). Task switching. *Trends in Cognitive Sciences* 7, 134–140.

- O'Connor, A. M., C. Bennett, D. Stacey, et al. 2007. Do patient decision aids meet effectiveness criteria of the international patient decision aid standards collaboration? A systematic review and meta-analysis. *Medical Decision Making*, 27, 554-574.
- O'Connor, A. m., Llewellyn-Thomas & Flood Ann, B. (2004). Modifying unwarranted variations in health care: Shared decision making using patient decision aids. *Health Affairs*, 63-72.
- Paulos, J. A. (1988). Innumeracy: mathematical illiteracy and its consequences. New York: Hill and Wang.
- Park, D. C. (2000). The basic mechanisms accounting for age-related decline in cognitive function. In D. C. Park & N. Schwatrz (eds.), *Cognitive Aging: A primer*, 3-21,. Philadelphia: Psychology Press.
- Park, D. C., & Kidder, D. P. (1996). Prospective memory and medication adherence. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications*, 369-390, Hillsdale, NJ: Erlbaum.
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N., Smith, A. D., & Smith, P. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, 17, 299-320.
- Peters, E., Finucane, M. L., MacGregor, D. G., & Slovic, P. (2000). Appendix C: The bearable lightness of aging: Judgment and decision processes in older adults. In P. C. Stern & L. L. Carstensen (Eds.), *The aging mind: Opportunities in cognitive research*. Washington, DC: National Academy Press.
- Peters, E., Slovic, P., Västfjäll, D., & Mertz, C. K. (2008). Intuitive numbers guide decisions. *Judgment and Decision Making*, 3, 619-635.
- Peters, E., Vastfjall, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy

- and decision making. *Association of Psychological Science*, 17, 407-413.
- Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In F. I. M. Craik (Ed.), *The handbook of aging and cognition* (2nd ed.), 1-90, Mahwah, NJ: Erlbaum.
- Raz, N., Briggs, S. D., Marks, W., & Acker, J. D. (1999). Age-related deficits in generation and manipulation of mental images: II. The role of dorsolateral prefrontal cortex. *Psychology and Aging*, 14, 436-444.
- Reyna, V. F., & Brainerd, C. J. (2007). The Importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences*, 17, 147 – 159.
- Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, 18, 89–107.
- Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *American Psychological Association*, 135, 943-973..
- Reuter-Lorenz, P. A., Marshuetz, C., Jonides, J., & Smith, E. E. (2001). Age differences in behaviour in behaviour and PET activation reveal differences in interference resolution in verbal working memory. *Journal of Cognitive Neuroscience*, 12, 188-196.
- Rothman, R. L., Montori, V. M., Cherrington, A., & Pigone, M. P. (2008). Perspective: The role of numeracy in healthcare. *Journal of Health Communication*, 13, 583–595.
- Schott, B. H. et al. (2005). Redefining implicit and explicit memory: the functional neuroanatomy of priming, remembering, and control of retrieval. *Proceedings of the National Academy of Sciences, USA*, 102, 1257-1262.

- Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, G. H. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, 127, 966-971.
- Sheridan, S. L., Harris, R. P., & Wolf, S. H. (2004). Shared decision making about screening and chemoprevention. *American Journal of Medicine*, 31, 56 – 67.
- Simon, D., Loh, A., & Harter, M. 2007. Measuring (shared) decision-making: A review of psychometric instruments. *Z. Arzt Fortbild Qualitat*, 101, 259-267.
- Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, 27, 279-306.
- Squire, L. R., Wixted, J. T. & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: a new perspective. *Nature Reviews Neuroscience*, 8, 872-883.
- Statistics Canada (2011). The *Canadian Population* in 2011: Age and sex. Catalogue no. 98-311-X2011001. Retrieved online (December 2, 2013) from www.statcan.gc.ca.
- Statistics Canada (2003). The *Canadian Population* in 2003: Literacy. Catalogue no. 92-3210-X2145301. Retrieved online (December 2, 2013) from www.statcan.gc.ca.
- Stevenson, F. A., K. Cox, N. Britten, & Y. Dundar, (2004). A systematic review of the research on communication between patients and health care professionals about medicines: The consequences for concordance. *Health Expectations*, 7, 235-245.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185, 1124-1131.
- U.S. Department of Health and Human Services. (2000). *Healthy people 2010: Understanding*

- and improving health* (2nd ed., Vol. 1). Washington, DC: Government Printing Office.
- Retrieved from <http://www.healthypeople.gov/document/HTML/Volume1/11HealthCom.htm>
- Wagner, A. D. & Gabrieli, J. D. (1998). On the relationship between recognition familiarity and perceptual fluency: evidence for distinct mnemonic processes. *Acta Psychologica*, 98, 211-230.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.
- Wechsler D. Wechsler Adult Intelligence Scale. 3rd ed. San Antonio, Tex: *Psychological Corp*; 1997.
- Woloshin, S., and L. M. Schwartz. 2002. Press releases: Translating research into news. *Journal of the American Medical Association*, 287. 2856-2858.
- Woloshin, S., Schwartz, L. M., Black, W. C., & Welch, H. G. (1999). Women's perceptions of breast cancer risk: How you ask matters. *Medical Decision Making*, 19, 221 – 229.
- Yonelinas, A. P. 2001. Components of episodic memory: the contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356, 1363-1374.
- Yonelinas, A. P. 2002. The nature of recollection and familiarity: a review of 30 years of research. *Journal of Memory and Language*, 46, 441-517.
- Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. In P. Sedlmeier & T. Bestch (Eds.), *Etc. frequency processing and cognition*, 21-36, New York, NY: Oxford University Press.

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, 98, 287 – 308.