Supporting Transfer Time Predictions in Medical Dispatch using Visualizations of Historical Data

by

Wayne Chi Wei Giang

A dissertation submitted in conformity with the requirements for the degree of Doctor of Philosophy

Department of Mechanical and Industrial Engineering
University of Toronto

Supporting transfer time predictions in medical dispatch using visualizations of historical data

Wayne Chi Wei Giang

Doctor of Philosophy

Department of Mechanical and Industrial Engineering
University of Toronto

2018

Abstract

Decision making in emergency medical dispatch is difficult due to the high time pressure and uncertainty faced by dispatchers. This is especially true in large-scale medical transport systems, such as Ornge in Ontario, because of the large geographical area serviced and the limited air and land ambulance resources available. Oversight agencies have highlighted the need to support patient transport time prediction for different transport options to facilitate better dispatch decisions. This dissertation proposes the use of visualizations of historical transport time data as a decision-aid to support predictions of patient transport time. Specifically, this dissertation examines whether visualizing the variability of historical data, which can help decision makers understand the uncertainty of the process, may improve transport time predictions.

Historical time predictions recorded by Ornge were statistically analyzed and two field studies were conducted at Ornge. Transport time predictions were found to be an important part of the dispatch process, but they are often underestimated, and that dispatchers do not explicitly incorporate uncertainty information in their predictions, possibly due to time constraints. Based on these results, an interface and workflow for a decision-aid visualizing historical data was proposed. In order to support the design of this decision-aid, two experimental studies were conducted to examine the influence of display format and context information on prediction behavior. The experiments were based on a proposed framework, informed by the literature, for how individuals use historical data visualizations to predict values of a variable. The experimental results provided evidence that both display format and context information impact prediction behavior, however the underlying process used by the participants appeared to differ

from that suggested in the proposed framework. Thus, further research is needed to improve this framework.

Overall, this dissertation adds to the very limited literature on supporting medical dispatch decisions through the design of a decision-aid for transport time prediction. In addition, this dissertation provides preliminary evidence for how individuals use visualizations of historical data to generate predictions of variables and what factors may influence these predictions.

Acknowledgments

This dissertation was a long journey and I was helped along by many more people than I can thank in these acknowledgements, so I just wanted to give my thanks and appreciation to everyone who has been part of my life over these past years for your loving support, friendship, and advice.

In particular, I want to thank my family. Mom and Dad, I would not be the person I am or dared the dreams I have dreamt with your support. Also, to my younger sister Amanda, who I now go to for advice and guidance, I always tell everyone that you're the real Professor Giang of the family.

I would also like to thank all the current and former members of the Human Factors and Applied Statistics Lab. I knew it was time to graduate when I was counting the number of cohorts of master's students that had graduated while I was doing my PhD rather than the years. Thank you for putting up with me all these years.

Thank you, Ellen and Olivier, for taking part in my defense committee. The insights you provided me have opened new avenues of research that I'm excited to pursue in the future!

To my committee, Mark and Greg, thank you for your guidance, support, and challenging me when I needed to be challenged (mostly about scope!). I know that I have become a better researcher, scholar, and member of the human factors community because of your advice.

Finally, to my advisors. Russell, thank you for all your help, advice, and expertise over these last few years. You always knew how to focus and ground my research and helped keep me on track. Birsen, I have learned so much over the course of my PhD and I'm sure I will even learn more form you in the future. I'm looking forward to putting these lessons into practice in the future to make you proud. Thank you.

Table of Contents

List of Tables

List of Figures

Chapter 1

1 Introduction

1.1 Time Estimation in Medical Dispatch

Cognitive aids, such as decision support tools, have been used to improve human decision making behavior in a variety of safety critical domains. These aids can supplement or replace parts of the decision making process that have traditionally been the responsibility of a human decision-maker. One domain where cognitive aids may be particularly useful in is Emergency Medical Dispatch (EMD; Fletcher & Bedwell, 2017). EMD refers to the management of emergency medical resources, such as paramedics and ambulances, in responding to requests for assistance (Clawson & Dernocoeur, 1998). Dispatch decision making can be difficult to support due to the complexity of the work environment (Chow & Vicente, 2002; Wong, Hayes, & Moore, 2005) and the information processing limitations of the decision makers (Palmer, 2017). Such decisions are made in highly dynamic situations characterized by limited and uncertain information and high time pressure. When confronted with these dynamic situations, decision makers in general, often make non-optimal decisions (Payne, Bettman, & Johnson, 1993), which can have drastic consequences in safety critical or resource-intensive scenarios.

In recognition of the pressures faced by EMD decision makers, better support for making medical dispatch decisions was one of the key recommendations of the Ontario Office of the Chief Coroner in a review of air ambulance related deaths in Ontario, Canada (Office of the Chief Coroner for Ontario, 2013). Specifically, the report highlighted the need to provide a list of possible transport options (e.g., fixed-wing aircraft, helicopter, and land ambulance transport options), along with expected transport times for each option, to key medical dispatch decision makers within Ornge, the primary provider of air medical transport in Ontario. In support of this recommendation, previous work has provided evidence that dispatchers often have difficulty producing accurate predictions of transfer times (Fatahi, Donmez, Ahghari, & MacDonald, 2012). In particular, dispatchers at Ornge have been observed to underestimate the time required for transport using helicopters as well as land ambulances. This underestimation can result in misallocation of resources and inappropriate dispatch decisions.

Time estimation is a cognitive process that has been found to be highly influenced by cognitive biases, such as the planning fallacy (Buehler, Griffin, & Peetz, 2010). Previous research supporting the planning fallacy (Buehler, Griffin, & Ross, 1994; Kahneman & Tversky, 1982b) has found that individuals tend to cope with uncertainty while making time estimations by adopting a singular "case-based" approach, by considering the specifics of the situation at hand. However, this approach may lead to time underestimates because these estimates are generated by "constructing plans and scenarios, with some allowance of safety margins for unseen contingencies" (Kahneman & Tversky, 1982b, p. 152), and anchoring biases can result in more optimistic estimates (Kahneman & Tversky, 1982b). To avoid such cognitive biases, a distributional approach to considering uncertainty, which relies on the underlying frequencies of the event being considered may be more appropriate. Evidence-based decision making, which refers to the use of objective data to support the expertise of one or more decision makers, may help decision makers adopt a distributional approach. This dissertation proposes the use of visualizations of historical transfer time distributions to support evidence-based decision making in medical dispatch. Specifically, this dissertation aims to improve transport time predictions through the creation of a decision support tool for Ornge's dispatchers that visualizes historical Ornge transport times. In addition, at a more theoretical level, this dissertation aims to contribute to the understanding of how individuals use visualizations of historical data to generate predictions of variables, such as time, within complex decision contexts.

1.2 Background and Scope of Dissertation

The use of evidence-based decision making has a long history in medicine (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996) and has begun to spread to other areas such as social and scientific policy (Dodge & Mandel, 2012), economics (Reiss, 2004), and business (Pfeffer & Sutton, 2006). Decisions made using this approach are concerned with predicting the outcome of future events in situations that are characterized by large degrees of complexity and uncertainty that cannot be deterministically modelled. The contrast between predictions based on expert (i.e., clinical) knowledge only and predictions made with objective evidence (i.e., statistical) has been well studied in medicine (Grove & Lloyd, 2006; Meehl, 1954). Statistical methods for helping predict future events tend to outperform clinical expertise because human decision makers can be unreliable and inconsistent in their decision processes and may sometimes resort to using inappropriate heuristics that bias their decisions (Dawes & Corrigan, 1974; Meehl, 1954). Expert decision makers, however, may have information and expertise available that is difficult to model or codify into statistical models (Meehl, 1954). Previous research has shown that when an aid is properly designed, the combination of human expertise with a computational model can better support judgment under uncertainty compared to either alone (Miller, Kirlik, Kosorukoff, & Tsai, 2008). In situations where there is an asymmetrical distribution of information between a model and a human, as is the case in medical dispatch (e.g., physicians have knowledge about the medical condition of the patient), joint human-automation decision systems can outperform the models alone (Yaniv & Hogarth, 1993). Therefore, this dissertation adopts the approach of combining human expertise in medical dispatch with decision-aids that present visualizations of historical data to improve decision making behavior. These visualizations are expected to help the decision maker better understand the type of outcomes that have occurred in the past, and act as a reminder of the varying situations that may exist in the future.

Emergency medical dispatch decision making can be categorized as short-term decision making. Short-term decision making deals with making judgments and predictions about specific future events (e.g., the current patient that requires transfer), with high time pressures often leading decision makers to rely on less information while also having fewer cognitive resources available to process the limited information available (Payne et al., 1993). One of the main challenges of effective evidence-based decision making in short-term, time-critical contexts is how well the decision makers are able to understand and use (or not use, as appropriate) the evidence that is available, given their limited time and cognitive resources. Therefore, this dissertation also explores how individuals interpret historical data visualizations to support their decision making.

As mentioned previously, the Ontario Office of the Chief Coroner recommended that expected transport times be provided for all patient transport options in order to facilitate better dispatch decisions at Ornge (Office of the Chief Coroner for Ontario, 2013). While previous work at Ornge has generated a decision support algorithm that provides point-estimates for patient transport times using historical data (Fatahi, 2013), it did not assess how the algorithm predictions compared to dispatcher predictions. This dissertation presents statistical analysis comparing the algorithm performance to dispatcher performance and shows that the algorithm can outperform dispatchers. However, as stated earlier, a joint human-automation decision system may outperform either alone, especially for Ornge's dispatchers who often have access to contextual information about the current transfer that is not available to the algorithm. Within the

Ornge context, historical transport duration data can reveal the underlying distribution of this variable of interest, including central tendency and dispersion (variability), providing to the dispatcher the range of possible values that the process generating the variable may produce (e.g., weather affecting medical transfer times). Previous research has shown that displaying uncertainty information in addition to a point-estimate of central tendency leads to better decision outcomes (Joslyn & Grounds, 2015; Joslyn, Nadav-Greenberg, & Nichols, 2009; Savelli & Joslyn, 2013). While previous studies have examined how paramedics (Harenčárová, 2015) and medical dispatchers (Wong & Blandford, 2002) cope with uncertainty during the course of their work, no research has examined how dispatchers understand and use uncertainty information to predict transport times. This dissertation addresses this research gap through two field studies conducted at Ornge.

While uncertainty visualizations have been applied to many contexts, such as categorical uncertainty (Bisantz, Marsiglio, & Munch, 2005; Bisantz et al., 2011; Finger & Bisantz, 2002; Neyedli, Hollands, & Jamieson, 2011), geographical/spatial uncertainty (Burton & Mccarley, 2017; Kirlik, 2007; Pugh, Wickens, Herdener, Clegg, & Smith, 2017; Ruginski et al., 2016), and displaying meta-information (Bisantz et al., 2009, 2014; Guarino, Pfautz, Cox, & Roth, 2009), little research has focused on supporting the prediction of a variable from a historical data distribution. That is, earlier research focused on different types of tasks, e.g., predictions based on historical trends (e.g., hurricane trajectory) or classification of a system state (e.g., is the target friend or foe?), than the one found in medical dispatch. This dissertation addresses this research gap by examining how commonly used historical data visualizations (e.g., boxplots) are interpreted by individuals to support the prediction of future values of a variable, in particular patient transport times.

Overall, this dissertation deals with three major bodies of literature (medical dispatch decision making, time estimation, and uncertainty visualization), and focuses on a specific set of problems within these large bodies of literature. Within the medical dispatch decision making literature, this dissertation focuses on supporting dispatchers in making more accurate transport time predictions for different transport options. Within the time estimation literature (i.e., how people estimate time), this dissertation focuses on time-estimation support through decision aids. Finally, within the uncertainty visualization literature, the focus is on commonly used visualizations for data distributions for a single random variable (i.e., transport time).

1.3 Research Questions

This dissertation examines the following research questions:

- Practical Research Questions: What is the role of patient transport time predictions in medical dispatch, and can information about the variability of historical transport times improve these predictions?

- Theoretical Research Questions: How are commonly used historical data visualizations (e.g., boxplots) interpreted by individuals to support the prediction of future values of a variable, and what factors influence these predictions?

The first set of research questions was addressed using a statistical analysis of historical transport time predictions produced by Ornge's dispatchers and two field studies conducted at Ornge. The second set of research questions was addressed through two experimental studies that were conducted online with non-dispatcher participants.

1.4 Organization of Dissertation

This dissertation is organized into three sections: Understanding why and how to support the generation of transport time predictions in medical dispatch; Empirical studies on time prediction with visualizations of historical data; and Conclusions. The first section deals with the practical research questions listed above, while the second section deals with the theoretical questions. The chapter organization is as follows:

Understanding Why and How to Support Time Prediction in Medical Dispatch:

- Chapter 2 motivates the study of supporting time predictions in the emergency medical dispatch context. This chapter provides an introduction to interfacility medical transfers completed by Ornge, Ontario's large-scale air and land medical transport service. An analysis of the accuracy of transport time predictions generated by dispatchers compared to predictions generated by an algorithm (built on historical data) is presented. The major contribution of this chapter is evidence that historical data can be used in an algorithm to produce prediction with less error than dispatchers. However, a joint human-automation decision system may further improve transport time predictions. Portions of this chapter were published in Giang et al. (2014) in *IEEE Transactions on Human-Machine Systems*.

- Chapter 3 presents two field studies with medical dispatchers at Ornge that examine the medical dispatch decision process; in particular, how dispatchers think about uncertainty and variability when generating transport time predictions. The major contribution of this chapter is evidence that uncertainty about transport times is not explicitly considered by dispatchers, even though transport time prediction is a key component of dispatch decision making. Furthermore, two categories of contextual factors considered by medical dispatchers that may influence transport time predictions are identified. Portions of this chapter were published in Giang et al. (2015), in *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, and Giang et al. (2016), in *Proceedings of the 60th Annual Meeting of the Human Factors and Ergonomics Society*.

- Chapter 4 proposes a decision support tool for Ornge that incorporates visualizations of historical data to support the generation of transport time predictions. The tool is currently being implemented by Ornge. The evaluation of its effectiveness in operation is therefore left for future research.

Empirical Studies of Time Prediction with Visualizations of Historical Data:

- Chapter 5 presents background literature and relevant work on the use of visualizations of historical data and uncertainty visualizations as decision support. Based on this literature, a framework is proposed for how individuals predict variables, such as time, using visualizations of historical data. The framework covers the effects of display format, contextual information, and individual difference factors and proposes that individuals first interpret visualizations to generate an internal probability model of the historical data and then select a prediction based on this internal model.

- Chapter 6 presents an empirical study that examined how different commonly-used visualizations influence time prediction behavior. Display format influences prediction behavior; more variability information leads to more deviations from the saliently presented central tendency point. However, even with no variability information, people do deviate from the point-estimate; e.g., by making conservative predictions, suggesting that prediction behavior is also tied to factors outside of the information presented in the visualization. Further, the presence of variability information allowed participants to adjust their confidence in their predictions, which may have beneficial effects on how these predictions are used.

- Chapter 7 presents a second empirical study that examined how the presence of contextual information, i.e., information that may be relevant to the situation in question but is not explicitly incorporated in the visualization, affects time prediction behavior and strategy. The contextual information tested was based on different types of contextual information observed in the field studies (i.e., those that change the likelihood of different outcomes, and those that change the consequence of different outcomes). The empirical study showed that users adjust their predictions when using historical data decision-aids in response to contextual information and that their prediction strategies differ based on the type of contextual information provided.

- Chapter 8 synthesizes the results of the experimental studies to draw conclusions about how individuals use historical data to generate predictions of a future time value and discusses limitations of the current work. The relevance of these findings to time predictions in medical dispatch and the decision support tool presented in Chapter 4 is discussed.

Conclusion:

- Chapter 9 describes the key findings and limitations of the dissertation. Future work evaluating the findings of the dissertation with Ornge dispatchers and the influence of individual difference factors on time prediction behaviors and strategies is suggested.

Chapter 2

2 Transfer Time Prediction in Interfacility Patient Transfers at Ornge

This chapter provides an introduction to interfacility medical transfers completed by Ornge, the medical transport service in the province of Ontario and motivates the problem of supporting time prediction in medical dispatch. An analysis of the accuracy of transport time predictions generated by dispatchers compared to predictions generated by an algorithm (built on historical data) is presented. The major contribution of this chapter is evidence that historical data can be used to produce transfer time predictions with less error than dispatchers. However, a joint human-automation decision system may further improve transport time predictions. Portions of this chapter were published in Giang et al. (2014) in *IEEE Transactions on Human-Machine Systems*.

2.1 Interfacility Medical Transport in Ontario, Canada

Interfacility medical transfers are one type of patient transport provided by medical transport systems, and refers to the transfer of patients between hospitals or other healthcare facilities (Ornge, 2017b). While often occurring with less uncertainty than other types of emergency medical responses (e.g., on-scene responses, where patients are taken from the scene to a healthcare facility), interfacility transfers are still time-critical processes that require tactical decision making due to the multiple transport options that are often available and the greater latitude in terms of the options for transporting the patient.

Interfacility transfers are a key component of improving patient care and reducing morbidity and mortality rates in the regionalized critical care model (Singh & MacDonald, 2009). In such a model, healthcare resources and expertise are concentrated in a small number of centers of excellence whose critical care and other specialized units receive patient transfers from a larger number of referral hospitals. There are many benefits to this model of healthcare, including improved patient care and reduced costs. However, one of the major hurdles is the transfer of patients between different facilities (Singh & MacDonald, 2009). Patients who are critically ill may face potential risks during the transfer, and care is delayed until the patient arrives at the receiving facility. There is evidence that the benefits of having a regionalized care model outweigh the risks (Singh & MacDonald, 2009); having an efficient medical transfer system is one way of mitigating the risks associated with interfacility patient transportation.

In Ontario, Canada, interfacility patient transfers are largely handled by a single province-wide transportation medicine service, Ornge. As the sole provider for air and land critical care transport medicine services in Ontario, Ornge performed approximately 81,000 interfacility patient transfers and 7,000 on-scene responses in the five-year interval between 2007 and 2011. The majority of the interfacility transfers were emergent (an immediate threat to life: 42%) or urgent (stable but risk for deterioration or threat to life or limb: 21%). Ontario is a large province covering an area of roughly 1.1 million km^2 (424, 000 square miles) which is more than double the size of France, with the majority of its population of 13.6 million concentrated in the south, along the border with the United States. Due to the size of the area serviced and the distribution of the population, Ornge relies on a variety of vehicles including fixed-wing aircraft, helicopter, and land ambulances to complete interfacility transfers.

As of 2017 (Ornge, 2017a), Ornge has 12 bases located throughout the province (Figure 1) and a fleet of 11 helicopters (Augusta Westland AW-139s) and 8 fixed wing aircraft (Pilatus PC-12s), and a smaller fleet of approximately 13 critical-care land ambulances (Crestline Commander). Ornge also has a team of critical-care and advanced-care paramedics that allow it to service transfers and scene-calls that would normally not be possible with other paramedic crews. In addition to its own aircraft and personnel, Ornge also coordinates with a number of third party air medical carriers and local EMS that provide additional transfers when the patient does not require a critical-care team, or when Ornge is able to provide their critical- or advanced-care paramedics to support the transfer.

There are very few air ambulance services that service as diverse and large an area as Ornge, and with the level of expertise of Ornge’s paramedics and dispatch staff. Some comparable large-scale medical dispatch systems include the BC Ambulance Service in British Columba, Canada, and Air Methods in the United States. Most other ambulance services are responsible for only a few major hospitals and with pre-determined ambulance resources available to complete transfers between these facilities. In contrast, the proper management of resources and triaging of patient transfers requests across hundreds of hospitals is one of the unique challenges facing dispatchers of large medical transport systems such as Ornge. As healthcare systems are consolidated and become larger, and as regionalized-care systems become more widespread, the need for dispatchers to deal with difficult transport mode decisions will also increase.

Figure 1: Ornge Bases in Ontario, Canada

2.1.1 Dispatch Decisions and the Operations and Control Center

Within Ornge, dispatchers within the Operations and Control Center (OCC) are responsible for a number of decisions that impact the efficiency of transfers, including receiving and analyzing transfer requests, assigning proper medical personnel and equipment to these requests, and ensuring that resources are available when required. One such planning decision, that will be explored in more detail in this chapter, is the selection of the mode of transportation based on patient and transport factors. When a patient transfer is requested, there are occasionally multiple transport options available for transporting patients between the sending and receiving hospitals. These options include fixed-winged aircraft, helicopters, land ambulances, or a combination of them. The choice of vehicle and crew introduces a source of uncertainty in patient transfer times, but also provides experienced dispatchers with a method for improving patient care.

The dispatch team within the OCC is composed of different roles and responsibilities working in concert to complete medical transfers. Medical call takers receive transfer requests from hospitals, local EMS, and regional Central Ambulance Communications Centers (CACCs), and collect information about the patients. Transport planners plan and organize the logistics of moving the patient, which involve choosing methods for transportation (fixed wing, helicopter, and land ambulances) and coordinating between the various entities involved in the transfer (e.g., arranging for ground transportation for a patient arriving in a fixed-wing aircraft at an airport, and securing a helipad or other landing sites). The Transport Medicine Physician (TMP) reviews incoming transfers in terms of the level of care required and is responsible for medical triage decisions when multiple patients require transportation at the same time. Finally, the Operations Manager (OM) oversees the operation of the entire communications center and ensures that resources are being managed effectively at a strategic level. Figure 2 shows an overview of the dispatch process for a typical interfacility medical transfer.

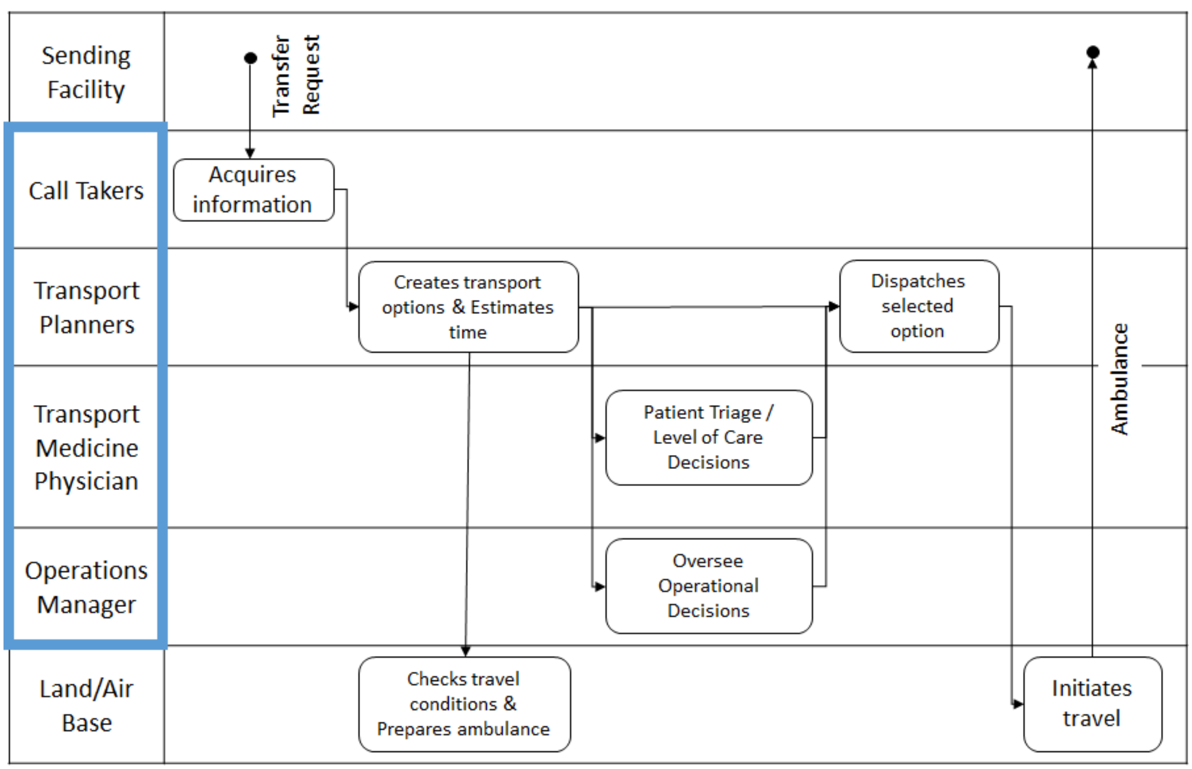


Figure 2: Ornge's dispatch process for an interfacility transfer (adapted from Fatahi, 2013)

This chapter focuses on the role of the transport planner who organizes the logistics of the transfer. Typically, transport planners are responsible for predicting transport times, as they are most intimately tied to the details of the transfer. The time predictions are then communicated to the other key decision makers (e.g., the OM and TMP) who use this information to help facilitate their dispatch decisions. Chapter 3 will provide further details about the goals and factors that the transport planners and other members of the dispatch team may consider during dispatch decisions.

2.2 Decision Support for Emergency Medical Dispatch

Transport mode decisions are often difficult because, at times, there are no clear advantages for choosing one vehicle over another. While air transportation is typically faster than ground transportation (Svenson, O'Connor, & Lindsay, 2006), a land ambulance may be able to more quickly respond and deliver a patient for short distance transfers due to the additional steps that air transfers may require (e.g., flight planning, pre-flight preparations, and additional land legs required for transferring between landing sites and hospitals). For emergent or urgent situations that have competing options for transportation, the air versus ground transport decision is a critical and time-sensitive choice. Thus, the choice of vehicle to use to service a call is often dependent on the predictions of how long the transfer will take.

Supporting transport mode decisions has not received much attention from the research community. Smith, Smith, Pletcher, Swope, and Kunst (1993) developed simple deterministic decision rules for trauma scene responses based on averages obtained from historical data from a single hospital. However, these decision rules were based on assumptions about the speed of different vehicles and did not consider factors that may only influence transfers to certain hospitals or along certain routes. Svenson, O'Connor, and Lindsay (2006) examined 145 cases in a comparison between air and ground transport times in interfacility medical transfers, and found that helicopter transport was always faster than ground transport. In contrast, Lerner, Billittier, Sikora, and Moscati (1999) generated a map for supporting transport mode decisions by showing zones where each type of vehicle was faster, and were able to identify regions where ground transportation resulted in lower out-of-hospital times than air. The map was generated by mapping and extrapolating historical transportation data on a Geographic Information System (GIS). However, the maps that were generated did not present information about the variability of out-of-hospital times for each region. Also, each of these studies only examined transfers from outlying hospitals and on-scene responses to a single trauma center and the generalizability of these findings for a small medical transport system to other systems and regions is unclear. As complex medical transport systems become more widespread, it would be important to consider transfers that happen across multiple sites and hospitals. Overall, there is still very little research in this area, especially regarding how medical dispatchers can make use of historical data to refine their transport decisions in time-critical situations.

2.3 Time Prediction in Medical Dispatch Decision Making

Time to definitive care, the time from when a call is received by dispatch to when the patient is finally handed off at the end of the transfer, is one of the important decision variables considered by dispatchers when making patient transport decisions (Harrington, Connolly, Biffl, Majercik, & Cioffi, 2005). Other time intervals that are important in dispatch decision making include time to bedside and out of hospital time. Time to bedside refers to the amount of time it takes for a medical transport team to arrive at patient's bedside and time out of hospital is the time interval between when a patient leaves the sending hospital and when the patient arrives at the receiving hospital. Thus, there are many important time predictions that dispatchers must generate when considering a transport option, and these predictions are often generated under high time pressure and uncertainty.

Research has shown that high time pressure alters decision making processes (Maule, 1997), often generating poorer decisions, especially when the decision makers are required to acquire and integrate information from multiple sources, as is the case in predicting transfer times for patients who require immediate care. Dispatchers who must arrange interfacility transport for urgent or emergent patients often need to make these decisions within minutes of receiving the call. Time pressure may also change the affective state of decision makers, leading to changes in risk taking behavior (Maule, 1997). Even under no time pressure, decision makers often rely on heuristics and biases (Tversky & Kahneman, 1974), which may result in potentially inappropriate decisions; and the presence of time pressure can lead to an increased reliance on these strategies. For emergent and urgent interfacility transfers, transport time predictions and the subsequent transportation mode decisions are both made under high time pressure and can greatly influence patient safety outcomes. For example, discrepancies between predicted and actual times can lead to inappropriate vehicle selections. Even when the choice of vehicle is not influenced, inaccurate predictions can lead to misallocation of resources and scheduling issues. For example, previous research with emergency medical responders has found that on-scene paramedics tend to underestimate time of arrivals (Propp & Rosenberg, 1991; Slack, Koenig, & Bouley, 1995), and that accurate predictions would lead to different medical oversight interventions (Slack et al., 1995).

In addition to time pressure, dispatchers also deal with large amounts of uncertainty when making decisions. The interfacility transfer process has high temporal variability. For example, prior work has demonstrated that factors such as precipitation significantly influence patient transfer times (Giang, Donmez, Ahghari, & Macdonald, 2014). Furthermore, information about the patient, the vehicles, and the routes (e.g., weather or traffic) often comes from second-hand sources or is predicted based on the dispatcher's experiences which introduces additional uncertainty to dispatcher predictions. Dispatchers may integrate this contextual information into their predictions, however, this integration may potentially increase workload. Overall, the time pressure and the uncertainty that the dispatchers face make the task of time prediction difficult.

Estimation and prediction of time durations have been studied outside of the medical dispatch domain (e.g., Halkjelsvik & Jørgensen, 2012; Kanten, 2011; König, 2005; Roy & Christenfeld, 2007; Thomas & Handley, 2008). One of the most studied phenomenon with respect to time estimation is the planning fallacy. Planning fallacy refers to the optimistic prediction of an outcome, most commonly time but also for other variables such as cost, even though historical data suggest otherwise (Buehler et al., 2010). There are a variety of explanations for why individuals may tend to underestimate time durations, including a focus on the specific case at hand (Kahneman & Tversky, 1982b), biased memories of how long the event has taken in the past (Roy, Christenfeld, & McKenzie, 2005), and motivation (Buehler, Griffin, & MacDonald, 1997; Byram, 1997). However, studies have also found that overestimations of time durations also can occur, particularly when the overestimations can help reduce stress due to time pressure in a scheduling task (Burt & Kemp, 1994).

2.4 Previous work with Ornge

Previous work conducted with Ornge (Fatahi, 2013; Giang, Donmez, Fatahi, et al., 2014) found that dispatchers adopt varying strategies to predict patient transport times (i.e., time to definitive care, time to bedside, and out of hospital time). For example, some planners use web mapping services to predict land vehicle travel times, whereas others may depend on their own knowledge of the region. In general, the planners break down the transfer process into components and predict a time for each component; however, there is variability in the number of components used by different planners.

In operation, the process of completing an interfacility transfer differs based on whether the transport vehicle is an air (i.e., helicopter or fixed-wing aircraft) or a land ambulance. However, the major medical transport steps are similar between both modes of transportation. Through onsite observations at Ornge's OCC and through ride-outs with paramedics during transfers, Fatahi et al. (2012) identified the following major medical transport steps for interfacility transfers: 1) vehicle departs base, 2) vehicle arrives at pick-up site (for land vehicles: sending hospital; for aircraft: can be an airport, a helipad at the sending hospital, or a helipad at a nearby location), 3) paramedics arrive at the patient site, 4) paramedics depart with the patient, 5) vehicle departs pick-up site, 6) vehicle arrives at the destination site (for land vehicles: receiving hospital; for aircraft: can be an airport, a helipad at the receiving hospital, or a helipad at a nearby location), 7) transfer of care (or delivery of the patient). For air vehicles, if the receiving and/or sending hospitals do not have a helipad (for helicopter) or the landing site is an airport (for helicopter or fixed wing), additional local land ambulance transfers are conducted to deliver paramedics to the patient site and/or to deliver the patient to the vehicle or to the receiving hospital. Figure 3 depicts the intervals which arise from these steps.

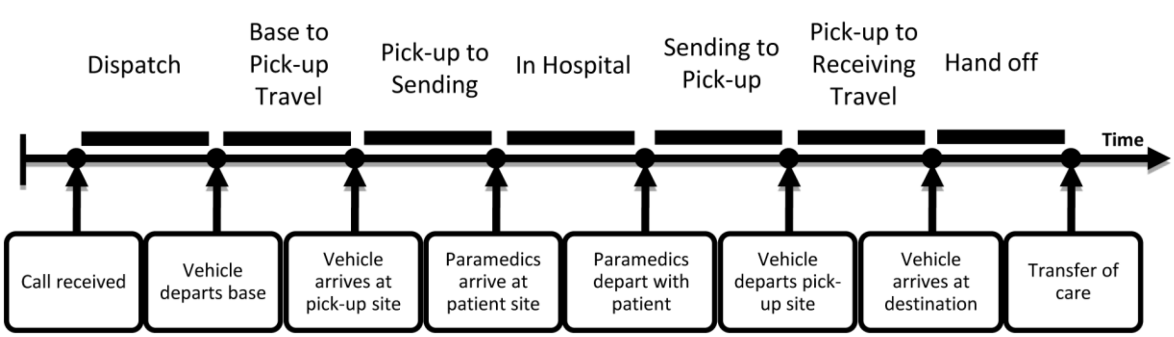


Figure 3: Major steps and time intervals of interfacility transfers

Since patient transfer time is a key decision variable in dispatch decisions, it was important to understand whether dispatchers could accurately predict these times for different types of vehicles. An initial analysis conducted by Fatahi et al. (2012) comparing actual transfer times to planner predictions found that planners tended to underestimate transfer times for both air and land transfers, but the underestimation was more prevalent for air transfers. In order to improve dispatcher predictions, a decision support algorithm was created to generate time predictions based on historical transfer time data (Fatahi, 2013; Giang, Donmez, Ahghari, et al., 2014). The algorithm breaks interfacility transfers into a number of intervals based on the major steps of the transfer process (Figure 3), and these intervals are used as the basic unit for predicting future transfers. Within each of the intervals, the subset of the historical data that share similar transfer characteristics for the interval (e.g., transfers between two hospitals that used the same vehicle) is selected. A point-estimate representing the average of similar past transfers is extracted using the median of this subset when there is enough historical data. When there is little or no historical data available, point-estimates are generated using other estimation methods including: 1) regression models for travel intervals (e.g., the flight or drive between the pick-up and drop-off sites), 2) real-time driving estimates from mapping software, and 3) a set of historical data of transfers that share some but not all of the same transfer characteristics as the patient transfer in question. The time to definitive care is calculated as the sum of all of the transfer intervals between when the call is received by Ornge and when care is transferred at the receiving hospital. A more thorough explanation of the algorithm can be found in Fatahi (2013) and Giang et al. (2014). Because the algorithm relies primarily on historical data, it is expected to account for nuances of specific transfer routes that the planners may be less familiar with (e.g., distance between helipad and hospital, differences between different facilities for in-hospital time). In the following section, dispatcher predictions of the time to definitive care are compared to those produced by this decision support algorithm.

2.5 Comparison of Dispatcher Predictions of Time to Definitive Care to Algorithm Predictions

As part of this dissertation, an analysis was conducted to compare how well dispatchers and the algorithm can predict time to definitive care to support decisions made within the dispatch interval indicated in Figure 3. The analysis used a stratified random sample of 171 interfacility transfers that occurred between 2010 and 2011. This dataset contained 77 (45%) transfers that included helicopter transportation, and 94 transfers (55%) that were completed entirely by land ambulances. The distances covered by the transfers ranged from 20 km to 806 km, with an average transfer distance of 130 km (SD = 84 km). Dispatcher predictions generated by the Ornge transport planners during the transfer process were recorded, along with actual transfer times during the transfer. In addition, the algorithm's time predictions were generated for each transfer.

As was found with the earlier analysis (Fatahi et al., 2012), dispatchers tended to underestimate the actual time to definitive care, and this tendency to underestimate was larger for air transfers than land transfers. Figure 4 (left) presents a scatterplot of dispatcher predictions versus actual transfer times. The 45 degree dotted line represents perfect predictions; data points below this line are underestimates and points above are overestimates. Overall, the magnitude of dispatch prediction errors (i.e., |predicted – actual|) was relatively large (mean, M = 53.56 min; standard error, SE = 3.58). Of the 171 cases, 74.3% were underestimates and 24.6% were overestimates (Figure 5, top). The percentage of over and underestimations differed significantly based on the mode of transportation with air transfers having a higher proportion of underestimations, $\chi^2(1) = 4.93$, p =.03. Furthermore, dispatcher prediction errors (i.e., predicted – actual) differed between air (M = -48.14 min, SE = 7.19) and land transfers (M = -26.70, SE = 6.00), t(169) = -2.31, p=.02. However, the variability of the air prediction errors (SD = 63.09) and land prediction errors (SD = 58.13) did not differ, F(76, 93) = 1.78, p =.45, ns.

The results of this analysis suggest that dispatchers were unable to accurately predict transfer times. More interestingly, there appeared to have a bias with underestimating air transfers to a higher degree compared to land transfers. This bias suggests that there may be more factors associated with air transfers that may not be considered during the prediction process. Air transfers often are more complex and require a larger number of steps (e.g., transferring the patient between the helipad and the hospital) that might be overlooked by dispatchers. There is also a general belief that air transfers are faster than land transfers as suggested by Svenson et al. (2006) and also brought up by Ornge staff. Thus, there is the possibility of such a belief influencing how dispatchers predict air transfers.

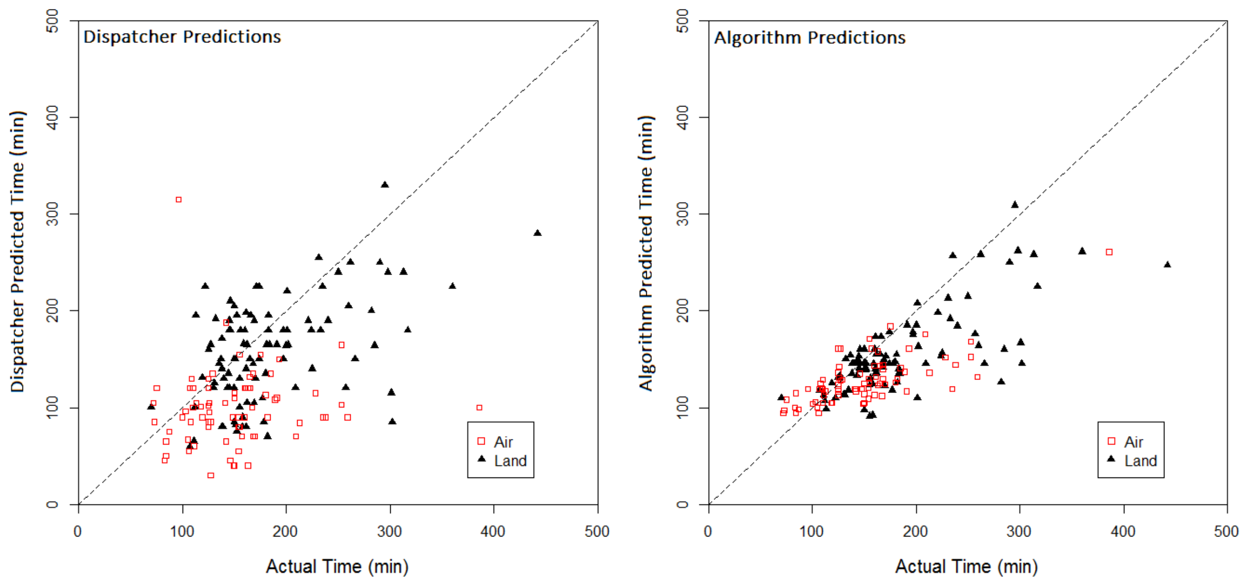


Figure 4: Time to definitive care: actual times vs. (left) planner predictions, and (right) algorithm predictions.

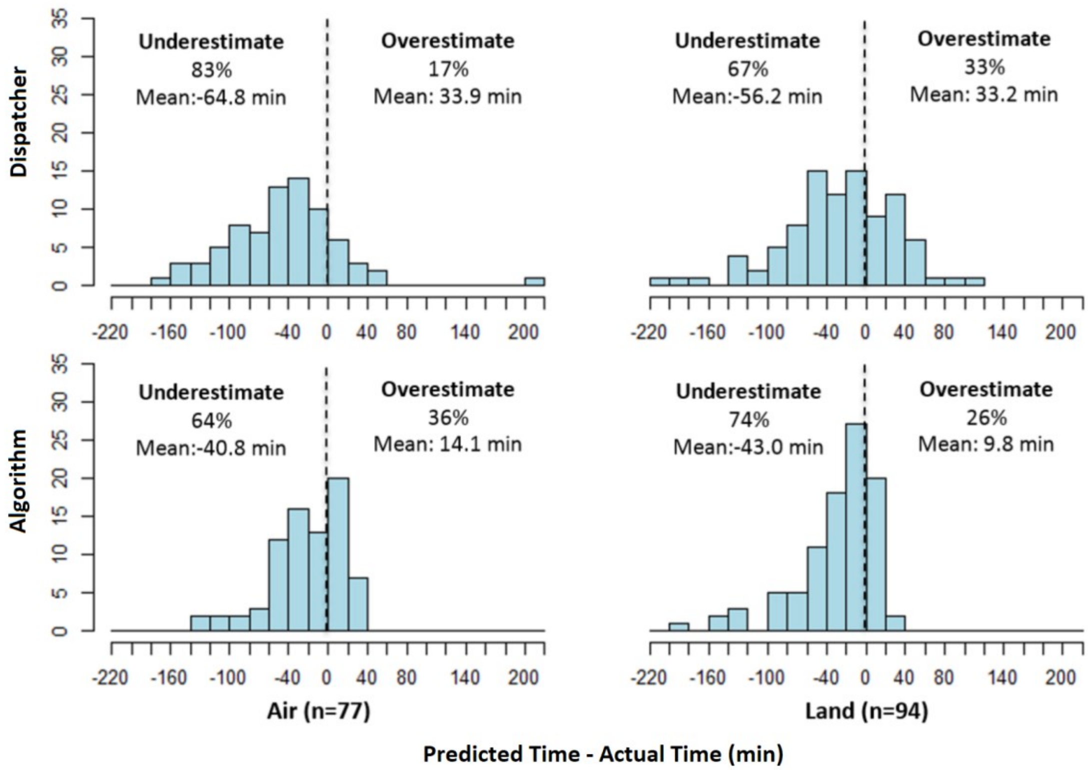


Figure 5: Histograms of prediction errors for air (left) and land (right) transfers for dispatcher (top) and algorithm (bottom) predictions.

Figure 4 (right) presents a scatterplot of the same stratified random sample of 171 interfacility transfers, and compares predictions generated by the algorithm with the actual transfer times. Again, the dotted line represents where a perfect prediction would fall. Similar to the dispatchers, the algorithm tended to underestimate transfer times, with 69.6% of the cases underestimated and 28.7% overestimated. Furthermore, for 74.4% of the cases, both the algorithm and the dispatchers had the same type of error (i.e., over or underestimate). However, Figure 4 also suggests that the algorithm prediction errors were smaller than those of the dispatchers, as many of the data points are located closer to the 45 degree line. Overall, prediction errors (i.e., |predicted – actual|) were on the average 21 min smaller for the algorithm ($M = 33.0$, $SE = 2.64$) compared to the dispatchers ($M = 53.6$, $SE = 3.58$), $t(170)= -7.62$, $p < .001$. This difference is practically significant as 20 minutes is crucial in urgent patient care and could result in an incorrect dispatch decision.

Unlike the dispatchers, the algorithm did not have a strong bias for underestimating air transfers (Figure 5). The proportion of over and underestimations did not differ between air and land transfers, $\chi^2(1) = 1.86$, $p = .17$, ns. In addition, the average prediction error (i.e., predicted – actual) for air transfers ($M = -20.88$, $SE = 4.23$) did not significantly differ from land transfers ($M = -29.50$, $SE = 4.39$), $t(169) = 1.40$, $p = .16$, ns. These results suggest that there is an opportunity to improve dispatcher predictions of the time to definitive care through the use of this and similar algorithms in combination with a human decision maker.

2.6 Improving Time Prediction by Supporting Human-Algorithm Collaboration

As a group, Ornge dispatchers have planned and carried out thousands of interfacility transfers and have developed experience and expertise in dealing with the difficult task of assigning resources to different transfers. However, with regards to transport time prediction, the algorithm analyzed in this chapter was able to outperform the dispatchers. This result agrees with the findings that statistical models tend to outperform expert decision makers (Dawes & Corrigan, 1974; Meehl, 1954). Even with the relatively simple methods used, the prediction errors observed with the algorithm were on average 21 minutes less compared to the dispatchers.

While it is tempting to remove dispatchers from the medical dispatch process, purely automated systems would not be appropriate as medical judgments often carry a degree of responsibility, and there may be social and legal issues with relying purely on statistical models for determining patient-related choices. Furthermore, the underlying models can fail to capture the constantly changing values and goals that exist in medical decision making; the improving or deteriorating medical condition of the patient may force the dispatchers to choose different transportation options. Brittleness, which refers to a lack of flexibility (e.g., a rigid set of variables considered by the algorithm from this chapter) when encountering unusual situations, is one weakness of automated systems, and is where a human decision maker can improve the overall decision making capabilities of the system (Cummings, How, Whitten, & Toupet, 2012). Finally, the results of the analysis reported in this chapter show that the algorithm predictions were still not completely accurate. Even with the creation of more accurate time prediction algorithms, there will always be some degree of error due to the uncertainty and inherent variability that exist in the world (e.g., traffic, weather). In fact, in all practical applications of decision making, there will always be some irreducible uncertainty that cannot be mitigated at the time of judgment (Hammond, 1996). Ornge dispatchers, while making less accurate predictions, are very adept at coping with uncertainty, and often take steps in mitigating or minimizing its effects on the overall transfer decision. For example, Ornge's transport medicine physicians come from emergency medicine backgrounds where acting with limited information and high uncertainty is part of the daily routine. Thus, the dispatchers can deal with situations which may not be captured by the point-estimates presented by a decision support algorithm.

Based on the literature presented in the introduction, this dissertation adopts the approach of combining human expertise with decision aids. For example, if information about the magnitude of the uncertainty associated with the algorithm prediction is available, it is possible that an expert dispatcher may be able to use this information (e.g., by changing the way they predict the transport time). One method of quantifying uncertainty in the transfer process is to examine the dispersion of historical transfer times that were used to create the algorithm prediction. Transfer intervals with low variability provide algorithm predictions that are fairly representative of that interval, while those with high variability deal with intervals that encompass a larger variety of situations. However, how dispatchers will use this uncertainty information is not well understood. In addition, some dispatchers may already be unconsciously accounting for uncertainty, for example by adopting more conservative decision criteria. To understand how dispatchers' predictions may be improved by the inclusion of evidence-based time prediction support, a more complete understanding is required regarding the medical dispatch process, in particular, how dispatchers think about uncertainty and variability when generating transport time predictions. These are the topics covered in the two field studies reported in Chapter 3.

Chapter 3

3 Interfacility Medical Dispatch Decisions

This chapter presents two field studies with medical dispatchers at Ornge that examine the medical dispatch decision process; in particular, how dispatchers think about uncertainty and variability when generating transport time predictions. While the original field studies were broader in scope, this chapter deals specifically with the two following goals: 1) understand if and how dispatchers use uncertainty information in transfer time prediction and 2) understand the major decision goals for Ornge's dispatchers in order to gain insight into which factors may impact transport time predictions. The major contribution of this chapter is evidence that uncertainty about transport times is not explicitly considered by dispatchers, even though transport time prediction is a key component of dispatch decision making. Furthermore, two categories of contextual factors considered by medical dispatchers that may influence transport time predictions are identified. Portions of this chapter were published in Giang et al. (2015), in *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, and Giang et al. (2016), in *Proceedings of the 60th Annual Meeting of the Human Factors and Ergonomics Society*.

3.1 The Role of Uncertainty in Dispatcher Time Predictions

As mentioned at the end of Chapter 2, this dissertation proposes the adoption of a joint human-automation decision system for the medical dispatch domain. As also presented earlier, the tools that have been developed so far present point estimates for their transfer predictions. The inclusion of uncertainty information associated with these point estimates can facilitate this joint human-automation decision system, as expert decision makers can leverage their prior knowledge and also bring in information about the current situation that may not be encoded within the automation. One way of quantifying uncertainty is through the transfer time variability recorded in historical data. To this end, the first field study explored whether and how dispatchers think about uncertainty while making transfer time predictions; and whether dispatchers thought that information about the variability in historical transfers would be useful in their decision making. This study was conducted over a period of three months from April to July in 2014.

3.1.1 Participants

Onsite observations were conducted with 4 Ornge transport planners. Two of the planners had less than one year of experience working as a transport planner, while the remaining two had more than 4 years of experience as transport planners and had also previously worked in other dispatch positions. One of the participants had a background in medicine; the other three participants had aviation backgrounds. The study was approved by the University of Toronto Office of Research Ethics. The relevant ethics documents for the two field studies are included in Appendix A.

3.1.2 Method and Procedure

A contextual inquiry approach (Beyer & Holtzblatt, 1998) was used for the observations and interviews with the planners. Observations were conducted during working shifts inside Ornge's OCC (Figure 6). Each observation occurred over a two-day period, and each day lasted an average of 5 hours. Overall, the data collection took approximately 40 hours. The days of the observations were selected to avoid operationally busy days (e.g., weekends and holidays) and observations were made during off-peak hours (between 10am-4pm) to minimize potential disruptions to the safety critical tasks performed by the dispatchers. Two observers were present during each observation.

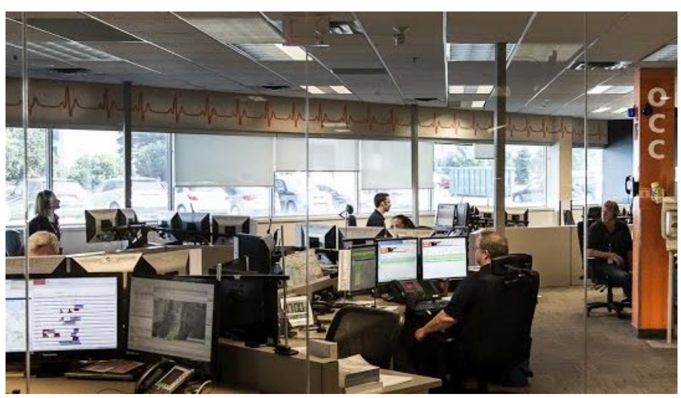


Figure 6: The Ornge Operations and Control Center (OCC) in 2014

The observations and interview questions mostly focused on urgent and emergent interfacility transfers. However, other dispatch calls, such as on-scene responses were also discussed with the planners. Throughout the observational period, the observers took notes on interactions with the computer-assisted dispatch (CAD) system and other tools used by the transport planners. After the completion of each transfer request, planners were interviewed about any elements of uncertainty that existed as part of their dispatch decision making. In particular, this study focused on the uncertainty of their predictions of the time to definitive care, which was defined to the planners as the spread of possible times that could occur due to unforeseen events or natural variation in the process. Planners were also queried about factors and contextual information that they incorporate into their decision making process which result in adjustments in their predicted times. As part of the observation process, transport planners were also introduced to the decision support algorithm described in Chapter 2, which provided point-estimates of the time to definitive care for the transfer requests they received. The following results were generated from a review of the notes from the observations and discussions between the interviewers.

3.1.3 Results and Discussion

Time prediction within the dispatch processes

High time pressure during dispatch decision making was observed during the contextual inquiry process. For example, medical triage decisions made by TMPs required the transport planner to produce, within minutes, multiple predictions of the transport time (i.e., for different transport options). However, one important finding from the contextual inquiry was that the prediction of transport times for different transport modes was only a small portion of the planner’s dispatch process, which included contacting multiple parties (e.g., pilots, paramedics, sending and receiving facilities, and local EMS services) and logging data into the CAD system. These other planning and coordination activities were the main focus of the planners’ work rather than the air versus ground transport mode decisions that were the focus of Chapter 2. Thus, the participants were unable to devote much time to predicting transfer times, even though it has been identified as a major area of improvement for preventing negative patient outcomes (Office of the Chief Coroner for Ontario, 2013). In the absence of the decision-aid, the dispatchers relied on a large variety of methods for generating predictions quickly. These included using their own experience and judgement, mapping software, and tables of travel times between specific locations.

Time pressure was still a major factor when the transport planners were provided with point-estimates generated by the algorithm; the participants had little time to reflect on these recommendations. Instead, they would often make use of the point-estimate immediately if they considered it to be reasonable, choose a transfer option (e.g., air vs. land) and would begin with the planning of the logistics of the transfer (i.e., asking the pilots to do a weather check). If they felt later that point-estimate could be more accurate (although it was reasonable), they would refine the predictions they entered into the CAD system. The use of a rough estimate to make the initial transport option decision was observed in general even without the algorithm predictions presented to the planners. One of the dispatchers stated that during busy periods, their main goal was to get resources out the door, and that estimated-time-of-arrivals could be updated en route if more accurate information became available.

Current and future use of uncertainty information

When asked about historical variability as a measure of uncertainty, participants stated that they would not intuitively think about uncertainty in this way. In fact, the planners did not appear to explicitly consider the uncertainty of predicted transport times. Instead, they tended to use best case scenarios or average times instead of considering a range of possible times or trying to anticipate rare events (e.g., mechanical problems with the vehicles). The planners stated that they favored generating predictions quickly, so that dispatch decisions could be expedited and they could focus on the more time-consuming tasks of organizing the logistics of the transfer. Historical transport time data, and the dispersion of these transfers, had not been available to planners in the past, and thus the interviewed planners were at a loss for how this information should be used to help with time predictions.

While uncertainty, as represented as the spread of possible transfer times, was not considered by the planners, the more experienced planners were interested in seeing information about the reliability of the predictions provided by a decision support algorithm as they were able to contrast their own experiences with the algorithm’s output. Therefore, reframing the variability of historical data as a measure of reliability, rather than uncertainty, may be beneficial as the variability can provide an indication of how consistent the transfer process is for a given set of vehicles, crew, and hospitals. In contrast, the less experienced dispatchers, were found to be more trusting of the algorithm’s predictions, since even when they felt the point-estimates provided by the algorithm may be incorrect, they were not as confident in their own predictions as the experienced planners. Therefore, reliability information can help build trust with experienced users, and can help less experienced users be more critical of a decision-aid's predictions.

The planners found the decision support algorithm useful for reminding them about the nuances of a particular transfer. For example, when the algorithm produced a longer than expected prediction for the helicopter to hospital portion of a transfer, one of the interviewed planners was reminded that the helipad was not a "walk-in", where they could place the patient on a stretcher and wheel them into the hospital. Instead a land ambulance must be booked to bring the patient from the helipad into the hospital, which takes longer than a simple "walk-in". This prompted the planner to take the required action of arranging for ground transportation. The presentation of historical dispersion information may also act as a similar prompt for the planners, as seeing larger variability in the historical data can remind planners that a particular transfer is more susceptible to disruptions, and this may result in adjustments in the dispatchers' prediction strategy.

Overall, the major findings of the first study were 1) the air vs. ground transport mode decisions are only a small portion of the overall dispatch task, leading planners to spend very little time generating predictions, thus 2) planners do not explicitly use uncertainty information as part of their time prediction process and mostly rely on best case or average times, but 3) planners do adjust their predictions to account for situational factors when they have time. While the transport planner's focus on organizing the logistics of a transfer may have limited the perceived benefit of uncertainty information, the study did identify potential benefits of presenting uncertainty information: assist with assessing the reliability of point-estimates provided by the decision-aid, and as prompt for planners to consider why a transfer may have large variability (i.e., more uncertainty).

3.2 Factors that Influence Dispatch Time Predictions

Since transport planners did not spend much time on predicting transport times, the second field study broadened the investigation to examine what decision goals the planners as well as other dispatchers, OMs and TMPs, focus on during their dispatch decisions. OMs and TMPs were included as they use planners' time predictions in higher-level dispatch decision making (i.e., resource allocation and patient triage), and uncertainty information related to these predictions may be more relevant for these roles. Further, as the first field study showed that planners appear to adjust their predictions based on the current situation (e.g., accounting for traffic and weather, their experiences with a particular hospital), the second field study also investigated the contextual factors that may influence time predictions.

3.2.1 Participants

Interviews were conducted with ten experienced Ornge dispatchers. Six were planners, 2 were TMPs, and 2 were OMs. All participants had been personally involved in managing interfacility and on-scene patient transfers in which they made critical decisions. Prior to their work at Ornge, the participants had a variety of backgrounds in operations, logistics, and dispatching in aviation, trucking, and the military. The study was carried out in two parts. Transport planners were interviewed during a period between September and December of 2015. The TMPs and OMs were interviewed between May and August of 2016. The study was approved by the University of Toronto Office of Research Ethics.

3.2.2 Method and Procedure

A methodology that has been previously used to study emergency medical dispatch (Wong et al., 1997) was adopted for the study which made use of Critical Decision Method (CDM) interviews. CDM is a retrospective knowledge elicitation method that is commonly used in naturalistic decision making contexts to understand how experts think during particularly difficult incidents using interviews where the experts reflect on a single previous incident (Klein, Calderwood, & MacGregor, 1989).

Interviews were conducted during working shifts inside the OCC. Participants responded to interview questions during downtimes between calls, and low-volume day shifts were selected to minimize the impact to the participants' main task of dispatching. During each interview, participants were asked to recall and 'walk-through' a single challenging decision making incident that involved an interfacility transfer or on-scene response in which they played a major decision making role. Alternatively, some interviews were conducted on an incident that occurred during the period of observation rather than a retrospective case. For the OMs, who oversee the operational decisions in the OCC, the critical incidents discussed involved multiple patient transfers or activities that involved preparing for transfers (i.e., shift changes and resource allocation).

From these incidents, a timeline of the incident was constructed jointly with the participants, and decision points were identified first by the interviewer and then verified by the interviewees. The decision points were elaborated on using a set of cognitive probes. The probes addressed the cues, tacit knowledge, experience, goals, and processes used during each decision point. After the interviews, the logs generated from the CAD system, which contained additional information about the incident (e.g., time stamps of major steps in the transfers), and the notes that were taken by the transport planner during the transfer were obtained.

For the six interviews with transport planners, a trained undergraduate research assistant conducted the interviews during working shifts inside the OCC. For the first three interviews, I was also present. Due to organizational constraints and research ethics, the interviews with the transport planners were not recorded. For the four interviews with TMPs and OMs, two trained undergraduate assistants and I were present. The interviews during this second phase of observations were audio recorded, due to amendments to the original ethics protocol. The audio recordings were transcribed by an undergraduate research assistant for analysis.

A structured approach (Wong, 2003) was used to produce an initial analysis of the CDM data. The steps of the structured approach consists of 1) creating a decisions chart (outlines the major decision points in the incident); 2) creating an incident summary (a written description of the incident that shows relationships between events that may not be clear in the decisions chart); 3) making a decisions analysis table; 4) identifying items of interest in each incident; 5) and comparing goals-states across multiple participants. In addition, short case studies were created from the interviews completed with the TMPs and OMs to describe major sources of uncertainty (Appendix B). The following results were generated from synthesis and discussions within the research team using the artifacts generated from the observation process (e.g., the structured analysis tables and the case studies).

3.2.3 Results and Discussion

Transport Planners

Transport planners appeared to have three major decision goals: maintaining situation awareness, matching the correct resource to the transfer, and planning the logistics of the transfer. Many of these goal states correspond to similar goal states that have been found previously in emergency medical dispatch (Wong, O'Hare, & Sallis, 1996):

1) *Maintain Situation Awareness (SA) – Global and Single Transfer:* Transport planners needed to have an awareness of where their resources were for two reasons: so that they could monitor ongoing transfers, which could take up to multiple hours to complete for long transfers, and to know what resources were available to use for future transfers. Thus, planners strived to maintain situation awareness of both the province-wide operational picture (global SA) and the specific details of a given transfer (single transfer SA). The planners have access to a number of tools (e.g., the CAD system, ACETech (used for tracking land ambulances; Ferno, 2018), Latitude (used for tracking air craft; Latitude Technologies, 2018)) that help them maintain both types of SA, and take time to update their awareness of their resources throughout the entire dispatch process.

2) *Match Resource to Transfer:* One of the major goals that the planners had to achieve when completing a transfer was finding the correct resource (e.g., the air vs. ground transport decision). Matching resources to transfers often required the creation of multistage 'transport solutions'. Transport solutions represented multistage plans for transporting a patient, often involving different vehicles, paramedic crews, and third-party resources (e.g., local EMS, third party air medical transport services). These solutions were created to satisfy three main outcome objectives: *provide the correct level of care, provide the fastest care possible, and provide the most cost and time efficient resource for maintaining Ornge's operational flexibility*. Overall, the planners attempted to balance each of the three outcome objectives as they created transport solutions. The relative priorities of the objectives differed based on the characteristics of the transfer (e.g., the patient's condition, the logistics of getting to the sending and receiving facilities, and weather) as well as the planners' SA of Ornge's resources at the time of the call. The planners would constrain certain parts of the transport while compromising on the other objectives. For example, patients who required the highest level of care would have a transport solution created that was constrained around getting a critical-care paramedic team onboard the vehicle, while less urgent patients would have transport solutions that could be optimized around minimizing costs and maximizing transport efficiency.

3) *Plan Logistics of Transfer:* Transport planners were also responsible for arranging the logistics of a transfer (e.g., calling ahead to make sure landing sites are secured and cleared, and arranging local ground transportation before an aircraft landed). By matching the needs of the transfer with the resources that could complete it, many of the logistical requirements are specified. However, planners must then take steps to ensure that the transfers can be carried out as planned, and must continue to monitor and adjust the transfer even after the resources had been dispatched. One such task was providing estimated times of arrival to the various parties involved in the transfer. The planner would use these estimated times to book the landing sites and local ground transport, inform the sending and receiving facilities about when to expect Ornge's paramedics, and to coordinate between the different vehicles used in the transfer.

Operations Managers and Transport Medicine Physicians

The OM and TMP roles dealt with different but related goals in comparison to the planners. One of the major difficulties for the planners was the balancing of the three objectives (level of care, fastest care, most efficient resource) as they built transport solutions. The OMs and TMPs had goals that dealt directly with these objectives and their jobs were to help the planners resolve conflicts between the three objectives. The goal states that were identified for OMs and TMPs were: maintain coverage, use most effective resource, and collaborate with expertise:

1) *Maintain Coverage:* Many of the decisions made by the OM were related to ensuring correct staffing levels and resource allocations both throughout the entire province (i.e., calling in extra resources when paramedics or aircraft were not available or had reached the end of their duty time outside of their home base), and within the OCC itself. While some of the decisions observed dealt explicitly with this resource allocation issue (see Case 3 in Appendix B), it was also a decision goal that was always considered by the OMs when they provided oversight for the decisions made by the transport planners within their teams.

2) *Collaborate with Expertise:* The OMs and TMPs were the OCC dispatcher roles that dealt with uncertainty most often. A lack of information about the current situation was a common cause of uncertainty (i.e., waiting for more information to come in from a 911 call; see Case 2 in Appendix B). However, both the OMs and TMPs also recognized

when they did not have the knowledge base to make an informed decision, at which time they made decisions to seek out a decision maker with the correct knowledge. This decision goal is explicitly built into the organization structure of the OCC with medical decisions being the responsibility of the TMP and operational decisions being the responsibility of the OM, a distribution of responsibilities that is common in air medical transport systems (Martin, 2006). The transfer decisions had both medical and operational characteristics and required collaboration between the TMP and OM. However, for this collaboration to happen, either the TMP or the OM must first recognize their own lack of knowledge and then explicitly reach out and work with the other member of the dispatch team. This also occurred when a TMP felt that the medical situation was outside of their area of expertise (see Case 1 in Appendix B where the TMP contacts a pediatric TMP due to the transfer of a patient going into labor).

3) *Use Most Effective Resource:* Both the OMs and TMPs were tasked for ensuring that the most effective resource was being used for each transfer, and these corresponded to the three outcome objectives mentioned previously for transport planners: *appropriate level of care*, *fastest care possible,* and *most cost and time efficient transfer*. For the TMP, ensuring the most effective resource manifested in the form of level of care decisions. Each transfer request was assigned a required level of care (primary care, advanced care, or critical care) that dictated which paramedic crews could be assigned to the patient. Ornge uses an algorithm to automatically assign a level of care for each new transfer, but these assignments could be overruled by the TMP. These re-assessments occasionally occurred when transport planners or OMs were considering alternative transport options that required more flexibility in terms of the paramedic crews. The TMPs would then need to balance the level of care they believe is required versus the constraints of current operational situation (i.e., what vehicles are available if the patient needed to be moved at that moment), which often depended on time to definitive care predictions provided by the planners for different transport modes. The OMs would also make similar decisions, but their focus was on the operational constraints (future coverage and cost).

Overall, it appears that: 1) many of the decisions made by Ornge are done as a collaboration between multiple decision makers who focus on different objectives that may conflict (i.e., medical considerations for the TMP vs. operational considerations for the OM), and 2) predicted transport times are a key decision variable in supporting many of the decision goals found for transport planners, OMs, and TMPs, and these predictions serve as an artifact for communicating information between the different dispatchers.

Ornge's transport planners appeared to have decision goals that focused on building transport solutions and planning the logistics of completing a transfer. The planners were able to quickly generate potential transport mode solutions (i.e., the multistage transport plans) to satisfy the requirements of the transfer. After developing such a plan, they would determine the feasibility of different plans by contacting the paramedics, vehicles, and hospitals involved. However, when multiple options appeared to be equally desirable or undesirable, sometimes due to the competing decision objectives (i.e., *appropriate level of care, fastest care possible, and most cost and time efficient transfer*), decisions about different elements of the transfer were delegated across the dispatch team (i.e., TMP and OM within the context of Ornge), as has been found in other studies of EMD (Furniss & Blandford, 2006). The planners would consult with the OM and TMP whose goal states dealt with resolving these conflicts. In contrast to the planners who focused on quickly recognizing a possible solution and then dealing with the logistics of the transfer, the OMs tended to consider multiple options, and would often ask their planners to produce alternative transport mode solutions (i.e., using a third party standing agreement carrier; see Case 4 in Appendix B).

The interviews also showed the importance of time predictions in Ornge's dispatch decision making. Predicted times were used in coordination activities performed by the planners in the form of Estimated Time of Arrivals (ETAs) that would be provided to Ornge's external stakeholders, as well as for planning of where and when vehicles should meet in multistage transfers. Time to definitive care predictions were given to the other decision makers involved in the transfer process (e.g., the TMP) to assist in their medical triage and level of care decisions. In choosing a transport option, the planners mainly relied on cues such as the distance of the transfer, their knowledge of the vehicle speeds, and their own experience in determining the closest or fastest resource. Only in cases that they felt were close did they explicitly calculate the time to definitive care.

In addition, the transport time predictions appeared to be one of the major artifacts for communicating information between different dispatchers in order to facilitate decision making.

This was especially true between the planners and the TMPs, who do not have the same level of operational understanding that planners have. When asked to make medical triage decisions, the TMPs would request and be provided with transport time predictions, and the planners (and OMs) would incorporate their knowledge of the current operational situation into their predicted times. Thus, the TMPs benefit from the contextual information about the situation that the transport planners use to adjust their predictions. These contextual factors will be discussed further in the following section.

3.3 Contextual Information in Medical Dispatch

One of the observations from the two field studies was that dispatchers considered many situation specific factors that influenced their understanding of how long it would take to transport a patient. Two types of contextual information were identified: likelihood information and consequence information. These factors were explored in the empirical study described in Chapter 7.

3.3.1 Likelihood-Information

Likelihood-information represent factors that should impact the transfer process, leading to changes in the expected transfer time. One example is with intubated patients, which often take longer to prep for transfer, leading to longer transfer times compared to non-intubated patients, if other factors are held the same. Intubation, and other patient related factors are just one type of likelihood-information that the dispatchers appeared to use to adjust their predictions, and these adjustments could be quite varied, as highlighted in the following quotes from the OMs:

> *"OM1: (the paramedic crew) could be there (the sending hospital) for 20 minutes or 3 hours, depending on what is wrong with the patient.*
>
> *Interviewer: I have seen planners who read through the patient logs and see if they are intubated and stuff.*
>
> *OM1: Yep, because we will look at things,* ***if a patient has any equipment lines and drugs so this person has cardiac monitor, oxygen they are on a stretcher and they are vented so they are not breathing on their own then they've got an arterial line and Endotracheal intubation, IV****,… so they have a lot of stuff*

> *hooked up to them. In sending facility all that equipment is hooked up to the hospitals equipment and our medics have to go in and disconnect each one of those things and hook it up to the transport version of the hospital equipment, make sure that the patient is stable and then take them out to the helicopter. That can take a couple of hours on some patients."*

> *"OM2:* ***Sometimes medics get there and the patient is not even remotely close to what they said they were, thank god, and things happen significantly faster. Other times, patients crash and unfortunately things take a lot longer****."*

Other examples of likelihood-information that were observed in the field studies include weather (i.e., an aircraft detouring around a storm front), traffic, paramedic crew experience and speed, and the location of the receiving units within a hospital.

3.3.2 Consequence-Information

In contrast, consequence-information deals with the gains or penalties (i.e., the value) for having certain outcomes, rather than the changes in the process. Sometimes different transfer times will result in different consequences for patient outcomes, the cost of the transfer, or resource management. For example, Case 1 in Appendix B, describes the decisions made by a TMP about the management and transport of a patient who was going into labor in a remote location. The TMP was actively trying to match the timeline of the labor with the timeline of the arrival of Ornge's paramedics (time to bedside) and when the patient could be transported to a receiving hospital (time to definitive care). As presented in the following quote, the TMP outlined possible consequences depending on where the patient delivered their baby; the consequence-information is highlighted in bold:

> *"TMP1: So if the patient is getting kind of further along in their labour you go, '… what's the worst possible scenario, is the worst possible scenario: stay in the nursing station, deliver in the nursing station or is it in fact take a chance to go and now the patient deliver in the aircraft'. And I think that we would generally agree that delivery in the aircraft is the least beneficial. So you go, 'ok',* ***you know the best would be to get to the receiving, the next best would be to stay in the nursing station, the worst would be to deliver in the aircraft****... so do you*

> *launch? and you go, 'well how far along is she, what is her history of labour, what is our crew configuration, how long is the trip, what's the weather' You know essentially you are trying to estimate what is the likelihood she is going to deliver in the aircraft. And you want to try and avoid that. By either delivering before or after."*

Other examples of consequence-information that were observed in the field studies include the flight crew's duty time limits (which can result in aircraft being stranded away from their home base), and paramedic and flight crew overtime. Ideally, consequence-information should not change time predictions for a given transfer.

3.4 Conclusion

This chapter described two field studies with actual Ornge dispatchers to better understand the dispatch decision making process. First, this chapter examined if and how dispatchers use uncertainty information during their transport time predictions. The results of the contextual inquiry observations showed that planners spent very little time generating time to definitive care predictions, even though these predicted times were a key component in many of the decision goals uncovered in the CDM interviews, suggesting that interventions (i.e., a decision-aid and/or simplifying the dispatch process) may help improve dispatch decisions. Furthermore, uncertainty was not explicitly considered by the planners, as they tended to rely on average or best-case scenarios rather than considering variability of historical transfer times. However, it appeared that the usage of a decision support tool by the planners may be improved through the presentation of uncertainty information: it can assist with assessing the reliability of point-estimates provided by the decision-aid, and as prompt for planners to consider why a transfer may have large variability (i.e., more uncertainty).

Secondly, the field studies provided insight on two categories of contextual factors that may be impacting the time predictions generated by Ornge dispatchers. Likelihood-information represents factors that should actually impact the transfer process, such as patient condition, weather, and the crew, and hence the transfer time. Consequence-information, on the other hand, represents factors that influence the consequences of a transfer, such as flight crew duty time limits and crew overtime but should not affect the predicted transfer time for a given option. Finally, the field studies also provided further evidence that dispatch decision making is a highly cooperative activity where transport time predictions act as one artifact for transmitting knowledge between dispatchers. These results provide a foundation for designing the workflow and interface of a decision support tool based on the algorithm described in Chapter 2, to help support transport time predictions. This design will be discussed in Chapter 4.

One limitation of the field studies presented in this chapter is that these studies were conducted at a single large-scale medical transportation system located in Ontario, Canada. As mentioned previously, there are few medical transport systems that service a geographical area and system of hospitals as large as Ornge does in Ontario, thus the generalizability of the findings to other medical transportation systems requires further investigation. For example, local EMS may not have access to the same paramedic resources and transportation options that are available to Ornge's dispatchers; thus, local EMS dispatchers may deal with other challenges to their dispatch decision making than those found for Ornge. However, large-scale medical transportation systems may become more prevalent as regionalized critical care models are adopted (Institute of Medicine, 2007) and hospital systems become consolidated (Cuellar & Gertler, 2003); the findings of this chapter can help understand the challenges faced by dispatchers within these systems. Furthermore, decision-aids are only one method for supporting the prediction of transfer times. For example, the introduction of new CAD systems to assist with the time consuming aspects of the transport planner tasks, such as entering and updating information within the CAD system, may allow dispatchers to spend more time on producing transfer time predictions and may also help improve dispatch decision making.

A second limitation of the field studies was the methods used in analyzing and interpreting the observation and artifacts generated (e.g., notes, transcripts, incident timelines, etc.) to generate the results. The analysis of the data largely consisted of synthesis and discussions within the research team, and thus may be influenced by the researchers' prior understanding of the domain and biases towards problems that were pertinent to transfer time prediction and uncertainty. While these topics helped focus and scope the field studies, the use of more rigorous methodologies such as thematic analysis (Braun & Clarke, 2006), grounded theory (Strauss & Corbin, 1994), and the use of multiple raters could help improve the reliability of the data analysis. However, the "quick and dirty" knowledge elicitation and analysis techniques used in this chapter have been found to be useful in supporting the design of visualizations and information technology for healthcare applications where there is limited access to decision makers and subject matter experts (Sockolow et al., 2017).

Chapter 4

4 Short-term Planning Tool for Ornge

Short-term planning (STP) refers to the logistical decisions that Ornge's dispatchers make for patients that require immediate interfacility transfer (i.e., urgent or emergent patients) and for scene-responses. In contrast to long-term planning, which often occurs a day or more in advance with the assistance of optimization decision support, STP decisions are made under highly dynamic situations characterized by limited and uncertain information and high time pressure. This chapter presents the proposed design for a STP tool interface and workflow that motivates the experimental tasks used in the following studies. The goal of the tool is to support STP decisions using the algorithm created in earlier work (Fatahi, 2013), which was briefly described in Chapter 2. Although this dissertation mainly focuses on interfacility transfers, the tool is also applicable to other types of transfers completed by Ornge, such as scene-responses. As mentioned earlier, this dissertation and hence the tool adopts the perspective that presenting variability information about historical transfers in addition to point-estimates can help better support dispatcher time predictions. First, the goals for the STP tool are outlined. Next, a short description of the workflow and interface of the current CAD system used by Ornge for planning the logistics of transfers is described. The proposed interface for the STP tool and how it fits in the existing workflow are then presented. This chapter was adapted from a proposal that was submitted to Ornge and the STP tool is currently being implemented by Ornge's developers as part of their new FlightVector CAD system. The evaluation of its effectiveness in operation is therefore left for future research.

4.1 Goals for the Short-term Planning Tool

Based on the literature cited earlier and the two field studies described in Chapter 3, four major goals for the STP tool were identified:

1. **Generate time predictions quickly**: Due to the high time pressure, and the limited amount of time available for generating time predictions, the tool must help dispatchers quickly predict transport times.

2. **Provide point-estimates**: Predicted transport times, such as the time to definitive care, time to bedside, and out-of-hospital time were found to be a key component of dispatch

decisions. These predicted times were an artifact used to communicate information between different members of the dispatch team and to external partners (i.e., hospitals or local paramedics). Due to the high time pressure, point-estimates are easier to communicate and use in subsequent decision making tasks by the dispatch team. Furthermore, point-estimates of predicted times are required for record keeping purposes within the CAD system.

3. **Encourage planners to choose a single best prediction**: The field studies found that planners adjust their predictions for the given situation based on their understanding of the current situation and their background knowledge. Because the predicted times are communicated to other dispatchers, the tool should encourage the planners to produce a single prediction that represents their best understanding of the current situation.

4. **Assist dispatchers with gauging the reliability of the suggestions**: Finally, the field studies showed that reliability information can help build trust with experienced users and can help less experienced users be more critical of a decision-aid's predictions.

4.2 Current Short-term Planning Interface in FlightVector

FlightVector is a commercially available dispatch software system that was customized by the vendor to suit Ornge's unique needs. FlightVector allows dispatchers to plan the logistics of an interfacility transfer or scene response, and these plans can incorporate multiple vehicles, paramedic and flight crews, and patients. This is accomplished through the "Plan" interface as shown in Figure 7. Through the Plan interface, the transport planner can build the transport solutions described in Chapter 3.

Within FlightVector, planners can assign vehicles and paramedic crews to patient transfers. The Plan window in Flight Vector allows the user to assign the waypoints associated with a transfer (i.e., the bases, airports/helipads, and hospitals). For a typical interfacility transfer the major waypoints would be the Ornge base where the vehicle is based, the sending hospital, and the receiving hospital. For more complex transfers, planners can input additional waypoints (e.g., to pick-up paramedics en-route to the patient). The Plan window also presents the estimated time en route (ETE) and estimated transfer times (ETT). The user enters their estimates for each portion of the transfer or relies on default values (e.g., for in-hospital times). As it is currently laid out, the Plan interface is built to help primarily with the logistics of the transport portions of a patient transfer (i.e., flight planning).

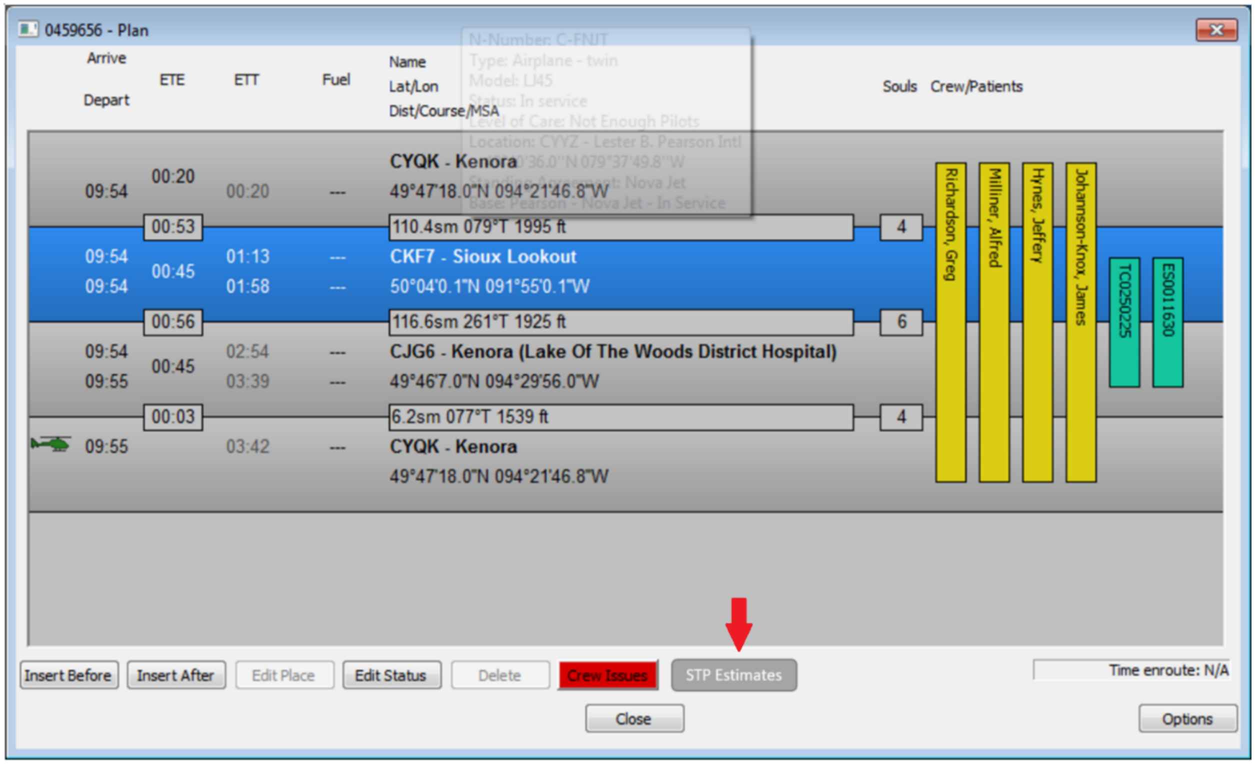


Figure 7: The Plan interface in FlightVector with proposed STP Estimates button indicated by the red arrow

A button ("STP Estimates") will be added to the plan screen that will start the STP tool. Information will be passed to the tool about the current plan, and once the user checks over the times predicted by the STP, the predictions will be passed back to the CAD system to automatically fill in the required time estimate fields for the plan, as shown in Figure 8. This automation will help reduce the workload of planners which may allow for more time to be devoted to the generation of predicted times.

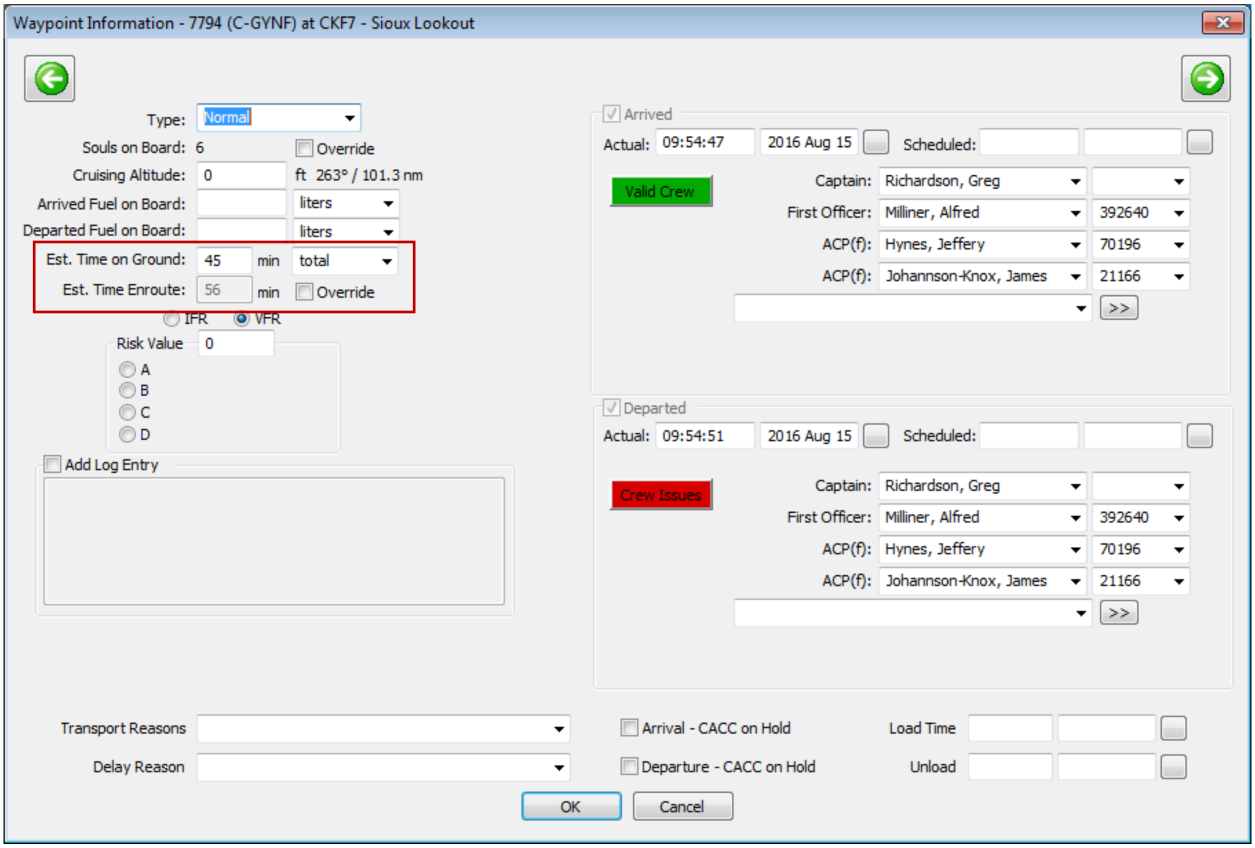


Figure 8: Waypoint interface from Flight Vector showing estimates for Time on Ground and Time En route highlighted in red

4.3 Short-term Planning Tool Interface

Once the user presses the STP Estimates button in FlightVector, they would be brought to the STP tool. The STP tool helps dispatchers estimate and predict patient transfer times by presenting to the user historical data and the decision support algorithm outputs. Dispatchers can either use the suggestions provided by the STP tool or they can adjust the transfer time predictions based on their knowledge of the current situation. In this section, the various elements of the interface are described in more detail.

The STP tool breaks down a transfer into the intervals presented in Chapter 2 (Figure 3), which represent the steps for interfacility transfers. Figure 9 shows the proposed STP tool interface, which is divided into two major sections: 1) Total Transfer Time Estimate (highlighted in red as an aid for the reader), and 2) Individual Interval Estimates. The Total Transfer Time Estimate section provides a quick overview of the entire trip duration, and provides the time to definitive care, the time to bedside, and out of hospital time. The total time to definitive care is represented as the sum of the time it takes to complete each interval of the transfer.

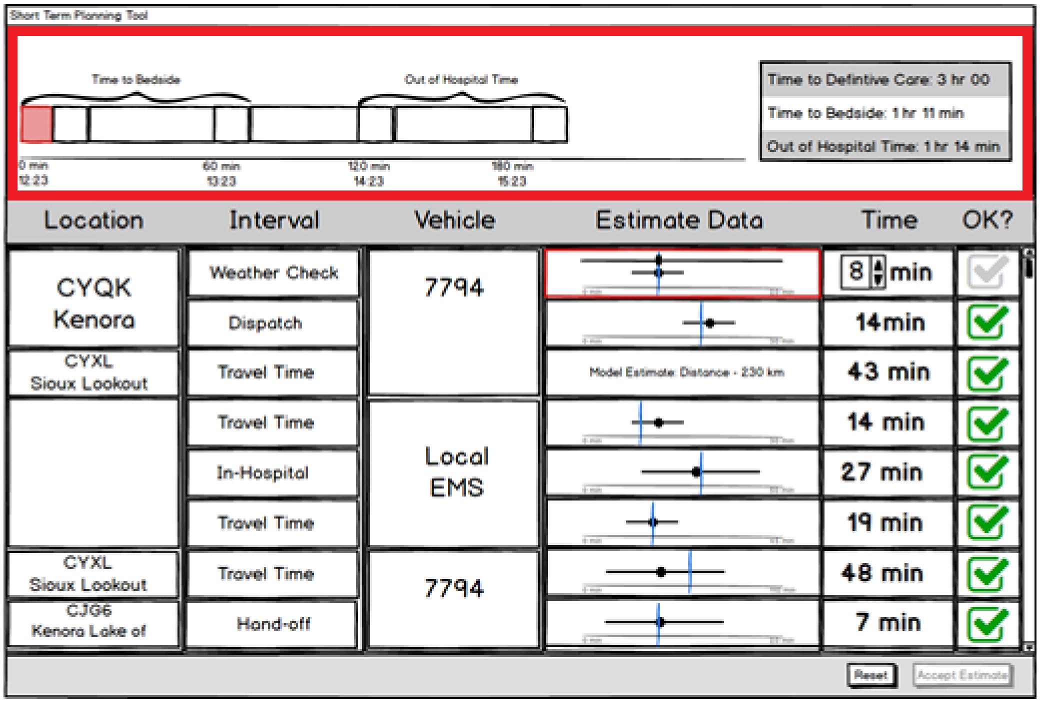


Figure 9: The STP Tool Interface, with the Total Transfer Time Estimate section at the top, and the Individual Interval Estimates section below

In the Individual Interval Estimates section, information is shown about each interval including the location, the vehicle used during that interval, the predicted time, and the data used to generate the estimate. The data shown under the estimate data column will depend on the algorithm used to generate the estimate (see Fatahi, 2013; Giang, Donmez, Fatahi, et al., 2014 for a description of the underlying algorithms). When historical data are used to predict, a visualization of the historical data is provided. Figure 9 shows one example of such a visualization using Mean and Standard Deviation plots. Other possible visualizations for displaying this information are discussed in Section 4.4 and will be covered in more detail in Chapter 6 of the dissertation. When models or mapping software are used to predict an interval, then the source of the prediction is clearly listed, to distinguish it from historical data, along with any other important information (e.g., the distance travelled). The last column (OK?) represents whether the user thinks the prediction is acceptable; all predictions are set to be acceptable by default and can be changed by the user.

If a user suspects a prediction is inaccurate and wishes to adjust the prediction they can click on the green checkmark, which unlocks interface controls to adjust the predicted value. When historical data are used for prediction, a slider appears on the visualization which allows the user to adjust the prediction to one that they think is more accurate. Users may also adjust their prediction by entering the value directly into the text input box in the time column, or by using the arrows to adjust the predictions up or down in minute increments. While this adjustment is being conducted, the interval that is being adjusted is highlighted in red in both the Total Transfer Time segment and the Individual Interval Estimate segments. The numeric prediction in the Total Transfer Time segment is also updated allowing users to see how their adjustments change the total time to definitive care. Once the user finds all the predictions acceptable, they can click on the accept predictions button, which passes these values back to the CAD system. Alternatively, they may hit the reset button to reset all predictions to the default tool predictions.

4.4 Historical Data Visualizations

Figure 9 utilizes Mean and Standard Deviation (Mean&SD) visualizations for historical data. Table 1 presents three commonly used methods for displaying historical data that can be used in the STP tool. The Mean&SD visualization shows the average historical time along with the variability of the historical data. The Boxplot visualization shows the minimum, 25th percentile, median, 75th percentile, and maximum values. Finally, the Dotplot visualization shows the historical data points and their distribution. The time scale of the historical data being displayed (e.g., the range of the x-axis of the visualizations) should be selected based on the entire set of historical transfer data for that interval in order to provide users with a sense of the range of possibilities. These display formats represent common methods for displaying and communicating data in scientific and technical fields. However, further research is required to determine how different visualizations influence prediction behavior. This question will be explored in the second half of this dissertation.

Table 1: Historical Data Visualizations

Mean & SD	[redacted]
Boxplot	[redacted]
Dotplot	[redacted]

4.5 Conclusion

This chapter proposed an interface and workflow for a STP tool that uses historical data and other methods to help support the patient transfer time predictions. The proposed STP tool was created based on design goals informed by the field studies from Chapter 3 and is currently being implemented at Ornge. However, a number of open research questions remain about how such a tool would be used by dispatchers, which are explored in the second half of the dissertation. The STP tool interface and workflow is one of the contributions of the dissertation. They also inspired the structure of the tasks used in the following experimental studies that examined how commonly used historical data visualizations (e.g., boxplots) are interpreted by individuals to support the prediction of future values of a variable, and what factors influence these predictions.

Chapter 5

5 Literature Review of Time Prediction using Visualizations of Historical Data

This chapter reviews the existing literature on how individuals generate predictions of future variables, such as time, when supported with visualizations of historical data. The major contribution of this chapter is a proposed framework for understanding prediction behavior that was used as the basis of experiments presented in the following chapters.

5.1 Prediction of Future Values under Uncertainty

Decision makers are often required to estimate current and future values of key decision making parameters and variables. These estimations occur across a wide variety of complexities and abstractions, from the generation of computational models and scenario analysis, such as in policy-making (National Research Council, 1996; Walker et al., 2003), to predictions of the future state of a single variable of interest, such as activity duration (Burt & Kemp, 1994; Koch & Kleinmann, 2002) or cost (Flyvbjerg, Holm, & Buhl, 2002; Khamooshi & Cioffi, 2013). This latter type of estimation is often found in short-term decisions where decision makers are required to estimate the outcome of a familiar but uncertain process. As presented earlier, in the medical dispatch context at Ornge, predicted patient transfer times are used to select an appropriate method of transportation, to coordinate patient hand-off between different vehicles, to support hospitals in preparing for incoming patients, and for record-keeping purposes for accountability reasons in order to justify dispatch decisions (Giang, Donmez, Fatahi, et al., 2014; Giang et al., 2016).

Decision-aids that present historical data on key decision making variables can be used to support grounded, evidence-based predictions of future values. Historical data can reveal the underlying distribution of the variable of interest, including central tendency and dispersion. The central tendency can be used in prediction for reducing long-term prediction errors (i.e., mean), or as an important indicator of the probability (i.e., 50% chance of the data being above or below the median) or of typical values (i.e., mode). Dispersion (variability) can provide insights into the uncertainties associated with the process that generates the variable of interest (e.g., weather affecting medical transfer times), with larger dispersion likely associated with higher process variability.

Uncertainty as a form of meta-information can be informative to decision makers (Bass, Baumgart, & Shepley, 2013; Bass & Pritchett, 2008; Baumgart, Bass, Voss, & Lyman, 2015; Bisantz et al., 2009; Lipshitz & Strauss, 1997; McQueary, Krause, Santos, Wang, & Zhao, 2004; Pfautz et al., 2006). Although previous studies have shown that displaying uncertainty information leads to better decisions (Bisantz et al., 2011; Joslyn & Grounds, 2015; Joslyn & LeClerc, 2012; Nadav-Greenberg & Joslyn, 2009; Savelli & Joslyn, 2013), these studies have focused mostly on decision outcomes rather than on the user predictions about a variable's outcome. Motivated by the medical dispatch decision context, where travel time predictions have significant value beyond their use in dispatch decision making, the focus of this dissertation is on variable prediction behavior as opposed to decision outcomes.

5.1.1 Predictions Aided by Decision Support

While extensive research has investigated how individuals predict the future (e.g., Burt & Kemp, 1994; Kahneman & Tversky, 1982a; Mannes & Moore, 2013), only a limited number of studies focused on how predictions of variables (i.e., time or location) are made when the user is provided with uncertainty information. Furthermore, some of these studies have focused on spatial predictions (Herdener, Wickens, Clegg, & Smith, 2015; Pugh et al., 2017), and have focused on extrapolations based on trend data, a task characteristically different than predicting a future variable value based on a historical distribution. However, research has shown that presenting information about the variability of a key decision cue can improve decision performance with a decision-aid beyond simply presenting information about the process that the automation is using to generate its' suggestions (Bass et al., 2013; Bass & Pritchett, 2008), highlighting the importance of showing historical variability information to support predictions.

Among those studies that focused on the prediction of a future value based on a distribution, Nadav-Greenberg and Joslyn (2009) and Savelli and Joslyn (2013) asked their participants to predict nighttime low temperatures based on forecasts with predictive intervals. This information was presented through numeric, verbal, and graphical display formats. Both studies found that participants tended to produce predictions that were lower than the point-estimate of the forecast (i.e., central tendency value) when a prediction interval was provided numerically or verbally, suggesting that participants adjusted their predictions in the presence of uncertainty information in these display formats. However, such a finding was not observed for the graphical display condition; participants' predictions were not significantly different than the central tendency value. A similar result was observed when participants were presented with just the central tendency value with no predictive interval; their predictions again did not differ from this central tendency value. These findings suggest that display format can have an effect on how decision makers use uncertainty information when making predictions. Ibrekk and Morgan (1987) investigated prediction behavior to compare nine different visualizations of snowfall forecast (e.g., violin plot, boxplot, cumulative distribution function) and found that participants attempted to locate the most saliently presented central tendency measure of the visualization as their best estimate of the forecast. It appears that with graphical displays, individuals are likely to rely on the saliently presented central tendency point to make their predictions.

In addition to considering the location of the predicted value relative to the central tendency point, prediction behavior may be characterized by the individual's judgement about how likely a predicted value is. Individuals have been shown to develop internal probability models to help translate graphical elements of uncertainty visualizations (i.e., uncertainty ranges or error bars) into probabilities (Tak, Toet, & van Erp, 2015; Tak et al., 2014). These internal probability models represent the individual's subjective interpretation of the probability distribution represented in the visualization. Tak et al. (2014, 2015) found that the internal probability models of uncertainty visualizations were best fitted by a normal distribution, suggesting that participants felt that the probability was highest toward the center of the visualization range, and decreased when it was further from the central tendency. However, the exact characteristics of the distribution varied based on the display format used and on the individual's numeracy. However, so far, no studies on prediction with uncertainty information have attempted to characterize prediction behavior using the decision maker's internal probability model.

The literature is also limited in terms of characterizing prediction behavior when uncertainty information is presented in different display formats. For example, individuals may rely on graphical features of historical data visualizations to assist with their predictions and/or they may rely on their internal probability models. Historical data are often summarized and presented in technical fields and scientific literature using graphical methods such as boxplots (Gschwandtnei, Bögl, Federico, & Miksch, 2016; Ibrekk & Morgan, 1987), and these visualizations vary in terms of how the uncertainty information is aggregated and which statistics are salient (e.g., mean vs. mode). Visualizations such as quantile dotplots (Kay, Kola, Hullman, & Munson, 2016) provide a potential understanding of the entire distribution. Other visualizations, such as error bars and predictive intervals (Savelli & Joslyn, 2013), aggregate uncertainty into bounds. Finally, visualizations that only show a point-estimate do not provide any uncertainty information. Since uncertainty information can provide the user with a better understanding of the underlying process they are trying to predict, the different methods for aggregating and presenting historical data may result in different prediction behavior.

5.1.2 Contextual Information

In addition to display format, there are many other factors that may influence how users interpret uncertainty information to predict future values. One such factor is contextual information. In situations where there is an asymmetrical distribution of information between a decision-aid and a human, as is the case in medical dispatch where dispatchers have information about the current situation that may not be explicitly encoded in an algorithm, joint human-automation decision systems have been shown to outperform the models alone (Yaniv & Hogarth, 1993). Contextual information can be the source of this information asymmetry. Previous work (Visschers, Meertens, Passchier, & de Vries, 2009; Wallsten, Fillenbaum, & Cox, 1986; Weber & Hilton, 1990; Windschitl & Weber, 1999) has found that individuals adjust their interpretations of both numeric and verbal representations of uncertainty based on contextual information.

One such contextual factor is the perceived base rate of the events being described by the uncertainty information, which was examined in a series of two experiments by Wallsten et al. (1986). In the first experiment, the authors examined how verbal indicators of probability, such as “likely” or “possible”, were interpreted by their participants in terms of numeric probabilities. Sixty meteorologists, professionals that are familiar with dealing with both numeric and verbal probabilities, were asked to report the numeric probability corresponding to a given verbal indicator of probability for four different types of medical scenarios. Two of the medical scenarios were selected by the experimenters to correspond to medical events that have a low base rate (i.e., is not likely to occur), and the remaining two were selected to correspond to medical events that have a higher base rate (i.e., is more likely to occur). In the second experiment, 72 undergraduate participants were asked to do a similar task with non-medical scenarios that were selected through a pilot study to have different base rates. Across both experiments, the authors found that the same verbal indicators of uncertainty (i.e., likely) were associated with higher numeric probabilities when the context referred to an event with higher base rate (e.g., "person will drop a non-required course after getting an **F** on their first exam" vs. "person will drop a non-required course after getting a **B** on their first exam". Thus, the individuals' interpretation of uncertainty information was influenced by their knowledge of the specific case in question, and Wallsten et al. (1986) proposed that the final interpretation was an averaging between the probability suggested by the uncertainty information and the base rate probabilities.

Similarly, research has also found that individuals consider the severity of a situation when asked to interpret verbal indicators of uncertainty. In a series of three experiments, Weber and Hilton (1990) found that in addition to base rate, event severity also changed the interpretations of verbal indicators of uncertainty. Each of the three studies were conducted with undergraduate participants. The first two studies used similar medical scenarios as the first experiment from Wallsten et al. (1986), which Weber and Hilton (1990) separated into high severity and low severity cases. The third study experimentally manipulated the scenarios using adjectives describing the condition (severity: mild vs. severe; base rate: common vs. rare; e.g., you will likely get a **severe** and **common** case of influenza vs. you will likely get a **mild** and **rare** case of influenza). Across the three experiments, participants were asked to assign a numeric probability to the verbal indicators of uncertainty (e.g., "possible", "slight chance"), and the results indicated that after controlling for base rate context effects, the severity of the scenario also changed the interpretation of the uncertainty information with higher severity scenarios being linked to higher numeric probabilities. While the base rate affects probabilities, the event severity should not.

The studies cited in the two paragraphs above highlight the fact that the usage of uncertainty information is dependent on context. However, these previous studies focused on uncertainty information dealing with membership within a class (i.e., the probability of having a disease or not), rather than the information about the variability of a stochastic process (i.e., a patient transport time). Whether context also plays a role in the interpretation of visualizations of variability information is an open question and is an important component in understanding how visualizations of historical data to support time prediction may be used in practice.

As reported in Chapter 3, two types of information that might influence dispatcher predictions were identified through field studies. The first type of contextual information, likelihood-information, is associated with the likelihood of certain outcomes (e.g., snow is likely to result in slower than normal transfer times). The second type, consequence-information, is associated with the gains or penalties related to certain outcomes (e.g., financial costs when the transfer time is long enough to warrant overtime pay). Likelihood-information comes from having knowledge that certain transfer related conditions are occurring, and may influence behavior similarly to the base rate context information examined by Wallsten et al. (1986). The likelihood-information may allow participants to estimate a conditional probability based on their understanding of the overall probability distribution. Individuals may adjust their predictions based on this posterior probability distribution, or adjustments may be made based on simple heuristics about the direction and magnitude of the effect (e.g., intubated patients take on average 10 minutes longer than normal). In each case, dispatchers that adjust their predictions based on likelihood-information may be able to predict transfer times that are tailored specifically to the situation, since the probability distributions for the time should be impacted by these types of factors.

In contrast, information about consequence (e.g., gains and penalties) may influence the utility of different predicted times. Changes in utility should guide decisions made using the time predictions but does not help with accuracy of time predictions. However, as seen with context information about severity (Weber & Hilton, 1990), consequence-information may still influence prediction behavior, since individuals can adjust their predictions in an effort to forestall bad outcomes, which is one strategy for coping with uncertainty (Lipshitz & Strauss, 1997). For example, dispatchers may overestimate the time required for a transfer in anticipation of unforeseen delays, or they may underestimate the time required if a longer transfer would require a costlier resource. If these systematic adjustments occur, then they may need to be mitigated using techniques such as training.

5.1.3 Numeracy and Education

Prediction strategies may also vary between different users due to individual differences, such as numeracy ability (Grounds & Joslyn, 2018; Tak et al., 2014, 2015). Numeracy refers to an aptitude and preference for using numeric information such as probabilities, ratios, and graphics, and has been found to influence the interpretation of uncertainty information. For example, Grounds and Joslyn (2018) found that, for most individuals, road-salting decisions that were made with numeric uncertainty information, in the form of probabilistic forecasts, resulted in better decisions than those made with only point-estimate forecasts. However, individuals with extremely low numeracy scores did not benefit from the uncertainty information. Higher numeracy has also been found to improve the interpretation of uncertainty visualizations (Grounds, Joslyn, & Otsuka, 2017; Tak et al., 2014, 2015). Tak et al. (2014, 2015) found that when asked to interpret visualizations of uncertainty, individuals tended to develop internal probability models that followed a normal distribution. The authors also found that individuals with lower numeracy generated models that were flatter and less normal (Tak et al., 2014).

The effect of education on the use of uncertainty information is less well understood. Ibrekk and Morgan (1987) did not find strong associations between their participants' level of education and their interpretations of uncertainty visualizations. However, Grounds and Joslyn (2018) found that participants with higher education made better decisions when provided with only point-estimate forecasts (i.e., no uncertainty information) than those with lower education, suggesting that higher education may help participants think probabilistically even when they are provided with no uncertainty information. Taken together, poor numeracy and lower levels of education may hinder the ability for individuals to interpret and use uncertainty information.

5.2 Proposed Framework for Predictions of Time using Visualizations of Historical Data

Based on the literature presented, a framework is proposed for how individuals use visualizations of historical data in order to predict future values of a variable, such as time (Figure 10). This framework is composed of five parts: the decision support, the context information, the individual difference factors, the individual's internal model, and the final predicted value. This framework posits that when individuals are presented with visualizations of historical data, they first internalize the information provided by generating an internal probability model. Using this probability model, individuals then choose a point that represents their "best estimate" of the variable they are predicting. As discussed in earlier sections, display format, contextual information, and individual factors can influence both the internal probability model that is generated and the point chosen from this model. The following sections further discuss the framework in terms of how display format and contextual information may influence prediction behavior; Chapters 6 and 7 present two experiments focusing on these aspects of the framework. The influence of individual difference factors, such as numeracy and education, are beyond the scope of this dissertation and is left to future work but was partially controlled for in the following experiments by screening participants using a test of statistical knowledge.

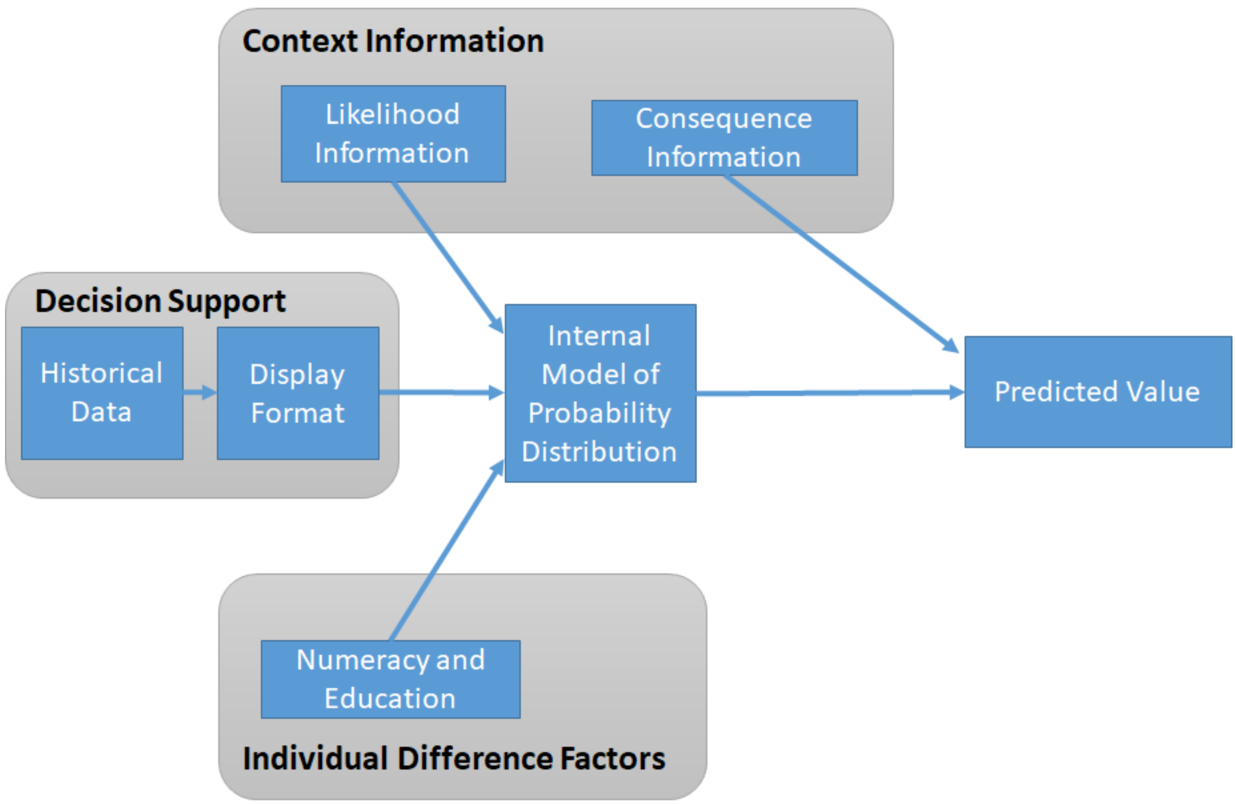


Figure 10: Proposed framework for time predictions with decision support

5.2.1 Prediction behavior based on display format

As mentioned previously, the graphical features of a data visualization can change its interpretation (Ibrekk & Morgan, 1987; Savelli & Joslyn, 2013; Tak et al., 2014, 2015), especially depending on how the uncertainty information is aggregated (Greis, Ohler, Henze, & Schmidt, 2015). While some studies have shown similar levels of performance with uncertainty visualizations based on different visual characteristics (Bisantz et al., 2005; Finger & Bisantz, 2002), these studies have typically provided the same amount of uncertainty information given that they were focusing on categorical membership probabilities. In the context of time predictions, commonly used display formats (e.g., Mean&SD, Boxplot) aggregate uncertainty to varying amounts (e.g., Mean&SD provides standard deviation bars whereas Boxplots provide both the interquartile range as well as the overall range).

Figure 11 provides an example of how display format may change a prediction made by a user of decision support. The display format used in this example is a dotplot where each observation is represented as a dot, and the visualization provides both a salient measure of the central tendency of the historical data (i.e., the mode, highlighted with an arrow in Figure 11), but also shape and skewness information that can be used to understand the underlying process. Based on the proposed framework, using this information, the user generates an internal probability model and then chooses a point along this model. With no context information, the user may choose the 50th percentile point as their prediction, as half of the historical data are above and the other half is below this point in the distribution. It is also possible that the user may rely heavily on the visualization, and choose a saliently presented central tendency measure, as was suggested by previous research (Ibrekk & Morgan, 1987; Nadav-Greenberg & Joslyn, 2009). In this case, different display formats may lead to variations in prediction behavior as they emphasize different central tendency measures. The influence of display format on prediction behavior was explored in an empirical study that is presented in Chapter 6.

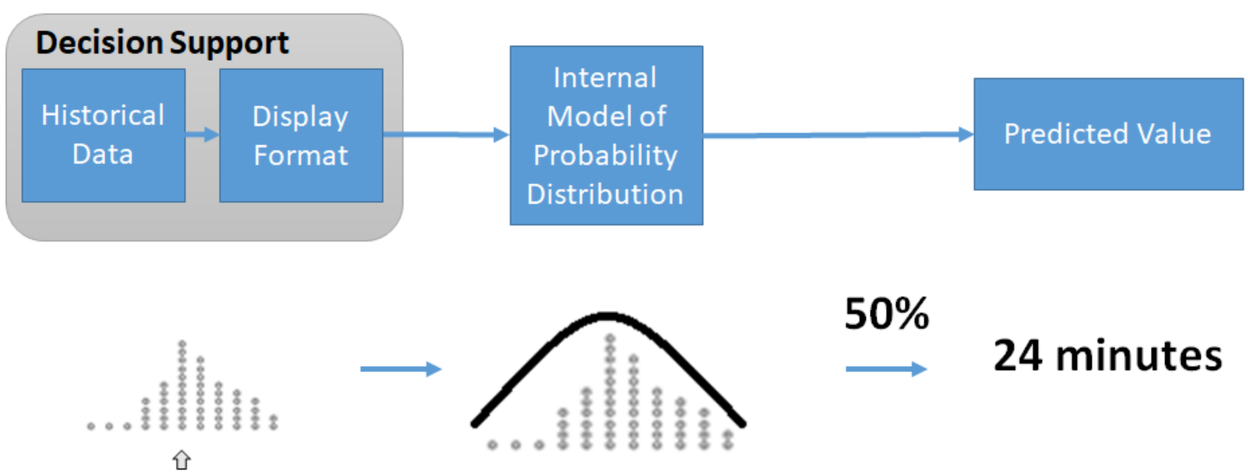


Figure 11: Example of the influence of display format on predicted value

5.2.2 Prediction behavior based on context information

Within the proposed framework, the historical data represent a set of similar historical events that the decision-aid has aggregated to help support the user's understanding of the process for which they are making a prediction. Context information, in this case, represents information that the user has access to about the current situation that may not be explicitly encoded into the historical data provided by the decision-aid. As mentioned previously, two types of context information that may be relevant to dispatcher's predictions of patient transfer times were identified: likelihood-information and consequence-information. Within the proposed framework, these two contextual information types influence different parts of the prediction process.

Figure 12 provides an example of how likelihood-information influences the predicted value. The framework proposes that the likelihood-information changes the internal probability model generated by the user. For example, information such as "this paramedic team is experienced and faster in preparing their patients for transfer" can change the user's interpretation of the historical data, resulting in a different internal probability model that better represents the subset of events that are directly relevant to a fast paramedic team. This would differ from the internal probability model generated without context information, as was shown earlier in Figure 11, and lead to a different predicted value.

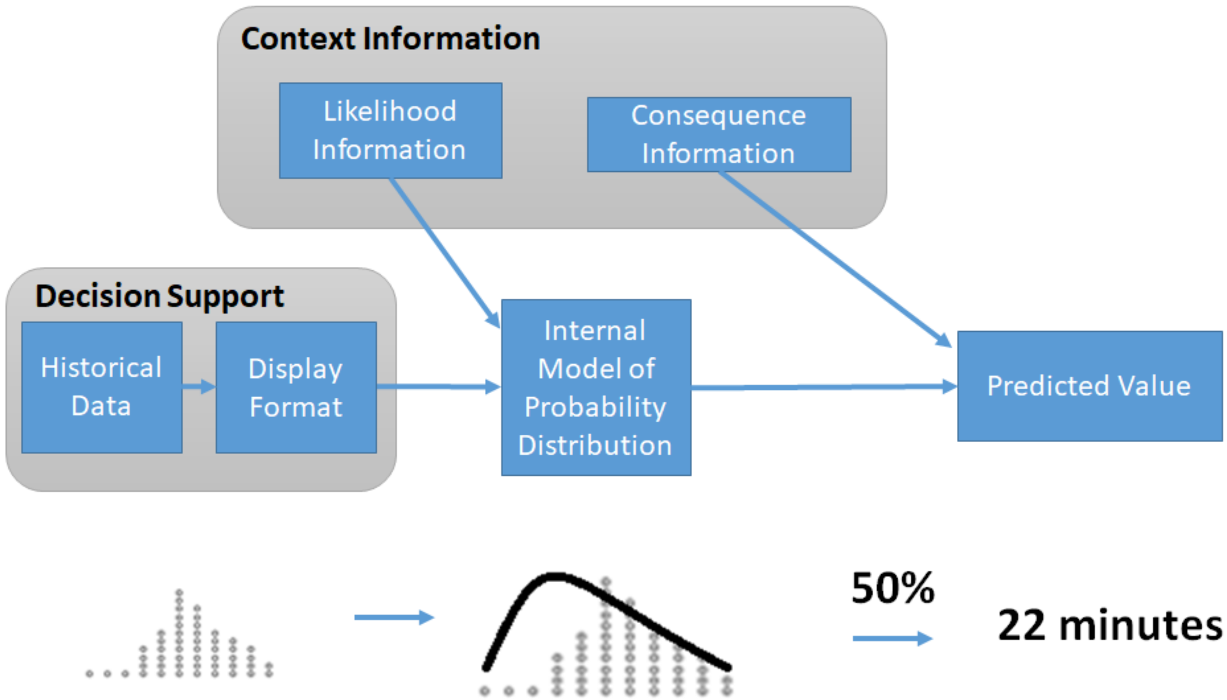


Figure 12: Example of the influence of likelihood-information on predicted value

In contrast, consequence-information, such as “if the transfer takes longer than 30 minutes, there is a higher chance of loss of life or limb for this patient” should not change the internal model generated but what the user selects as their prediction (Figure 13). Based on the severity of the consequences, users may select a more conservative value in order to avoid underestimating the time. However, this change in prediction strategy can be considered to be a bias if the goal is to minimize prediction error. The influence of context information on prediction behavior and strategies was explored in a second experiment. This experiment is presented in Chapter 7.

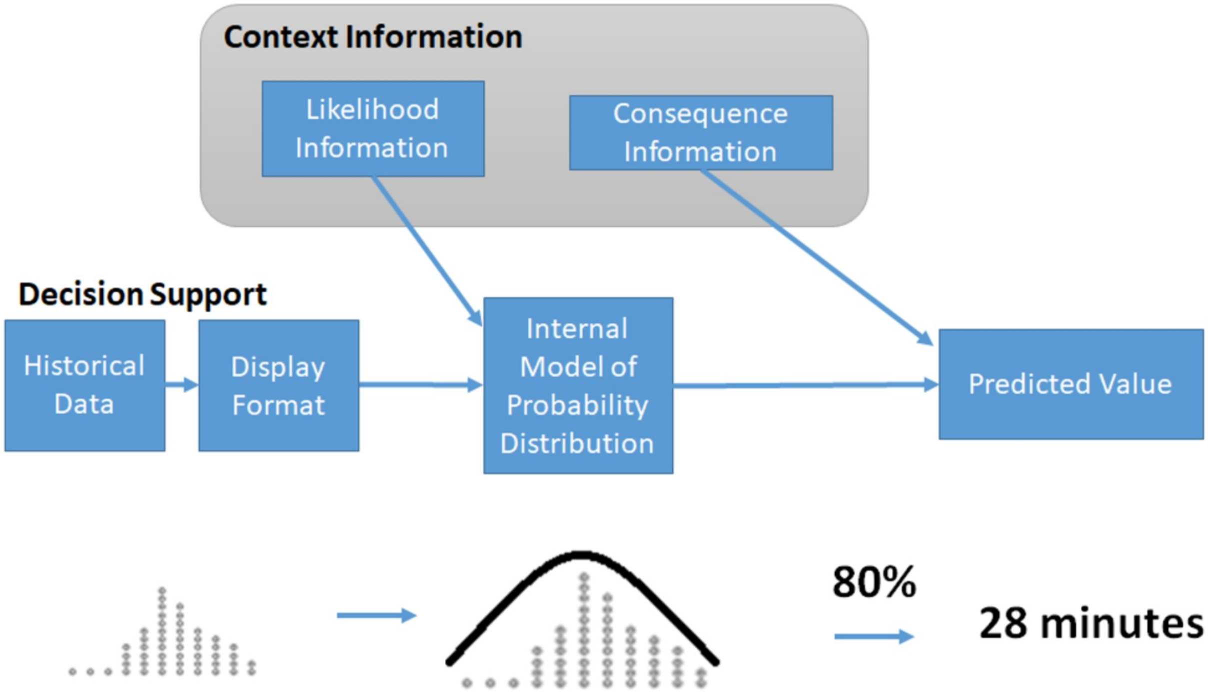


Figure 13: Example of the influence of consequence-information on the predicted value

Chapter 6

6 Influence of Display Format on Time Prediction

This chapter investigates how different historical data display formats influence the choice of a predicted value. In particular, this chapter examines whether, as suggested by pervious literature, saliently presented central tendency measures (i.e., median, mean, or mode) are likely to be chosen as the prediction. In addition, this chapter addresses the gaps in the literature identified earlier by examining (1) how predictions are related to the individual's internal probability model of the data, and (2) how the magnitude of variability in the process (i.e., how much variability there is in the historical data), along with the amount of variability information presented by a visualization (i.e., the display format), influence prediction behavior. Four commonly used methods for visualizing datasets containing repeated observations of a single process, such as the transport time between two hospitals, were evaluated: Median-only visualization, Mean&SD visualization, Boxplot, and Dotplot.

The major findings of this experiment were that the display format influences prediction behavior; more variability information leads to more deviations from the saliently presented central tendency point. However, even with no variability information, people do deviate from the point-estimate; e.g., by making conservative predictions suggesting that prediction behavior is also tied to factors outside of the information presented in the visualization. The presence of variability information, especially in the Boxplot and Dotplot visualizations, allowed participants to adjust their confidence in their predictions, which may have beneficial effects on subsequent decisions made using the predictions.

6.1 Experimental Scenario and Visualizations Tested

Due to the limited availability of the highly trained dispatchers at Ornge, an empirical study with non-dispatch participants was carried out to examine the influence of display format on prediction behavior. The study used a fictional scenario where participants were asked to predict the duration of scientific tasks completed by planetary exploration rovers using visualizations of historical task completion times. The unfamiliarity of this fictional scenario helped control for differences in background knowledge and encouraged participants to rely heavily on the visualizations, rather than relying on their own personal experiences and biases with the domain.

Four visualizations (Figure 14) were used, each with increasing amounts of information about the dispersion and shape of the distribution of historical observations. These visualizations were chosen from commonly used methods for displaying and communicating data in scientific and technical fields.

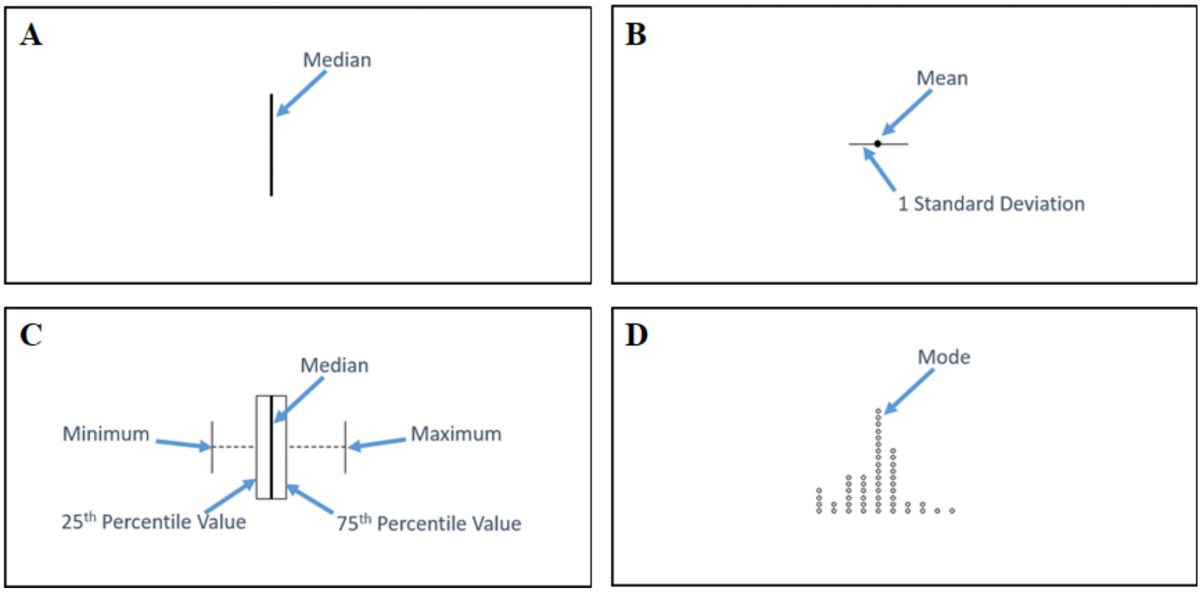


Figure 14: The visualizations of historical data used in the experiment, with the x-axis representing task duration: a) Median-only, b) Mean & Standard Deviation, c) Boxplot, and d) Dotplot. The arrows and text in blue present the additional information provided to the participants during training.

In the baseline, i.e., the central tendency only visualization (Median-only), the median of the historical sample was displayed as a vertical bar similar to the median line in a boxplot. The Mean & Standard Deviation (Mean&SD) visualization showed both a measure of central tendency and a symmetric measure of dispersion (i.e., the mean +/- one standard deviation). The Boxplot visualization provided a measure of central tendency (median) as well as two measures of dispersion (interquartile range and range). The Boxplot also provided information about the shape of the historical sample. The Dotplot provided shape information but used discrete markings for each observation, furthermore the mode of the historical data could be quickly identified using this visualization (the x-value that corresponds to the peak). Thus, each visualization had a central tendency measure that was saliently denoted; this central tendency measure was also highlighted during training as presented in Figure 14.

6.2 Methods

6.2.1 Experiment Design

The experiment was a mixed factorial design with visualization condition (Mean&SD, Boxplot, or Dotplot) as a between-subjects variable, and the presence of variability information (not present vs. present) and the variability magnitude (smaller vs. larger SD) as within-subjects variables (Table 2). Participants were randomly assigned to one of three visualization conditions (Mean&SD, Boxplot, or Dotplot), with 20 participants per condition. Each participant was presented with the Median-only visualization first as a baseline before completing their assigned visualizations (one of the three visualizations that provide variability information). Participants completed six trials with the Median-only visualization, followed by six trials with the other visualization type that they were assigned, resulting in 12 trials per participant. The latter half of the trials had both smaller and larger magnitudes of variability (i.e., smaller SD vs. larger SD); each variability level was repeated in three trials.

Table 2: Experimental design for display format experiment

	Variability Information		
Visualization Condition	***Not Present***; *trials 1-6*	***Present***; *trials 7-12*	
Mean & SD *participants # 1-20*	Median-only *datasets 1-6*	Mean & SD	Smaller SD *datasets 1-3*
			Larger SD *datasets 4-6*
Boxplot *participants # 21-40*	Median-only *datasets 1-6*	Boxplot	Smaller SD *datasets 1-3*
			Larger SD *datasets 4-6*
Dotplot *participants # 41-60*	Median-only *datasets 1-6*	Dotplot	Smaller SD *datasets 1-3*
			Larger SD *datasets 4-6*

Given that a visualization was repeated in multiple trials by a participant, different datasets were created to represent different rover task completion scenarios. Overall, six datasets were created to serve as the historical data (Appendix C); three had smaller SD and three had larger SD. The datasets were used twice, once for trials 1-6 and once for trials 7-12 (see Table 2). Each dataset consisted of 50 observations sampled from the normal distribution; the normal distribution was used as it represents many natural phenomena. Both skewness and kurtosis of these datasets were kept between -0.5 and 0.5 to ensure that the datasets appeared normal when visualized. Smaller SD datasets came from distributions with SDs of 2 minutes, the other three came from distributions with SDs of 5 minutes. Order of presentation of the datasets was randomized.

6.2.2 Participants

Participants were recruited using a paid screening questionnaire on Amazon Mechanical Turk (MTurk), an online platform for crowdsourcing tasks, which was open to all MTurk Workers in the United States and Canada with Masters Qualification, i.e., top workers as designated by MTurk who have a high degree of accuracy in their work. Participants were screened for self-reporting to have normal or corrected-to-normal vision and normal color perception. In addition, the screening questionnaire provided definitions and examples of central tendency (i.e., mean, median, and mode) and dispersion (i.e., range, interquartile range, and standard deviation) concepts. Participants were tested on these concepts and had to score at least 75% to qualify for the study. The experiment took approximately 1 hour, and participants were compensated a total of US$6.5 ($1.5 for the screening questionnaire and $5 for the actual experiment).

Sixty participants (37 male, 23 female, mean age=36.7 years, SD=9.3) completed the study; they had scored an average of 89% (SD = 11) on the screening test. Thirty-seven participants had taken at least one probability or statistics course at the post-secondary level and a further 12 had completed a probability or statistics course at the high school level. The study was approved by the University of Toronto Research Ethics Board. Informed consent was obtained from each participant for both the screening questionnaire and the actual experiment. The informed consent document as well as other relevant experimental materials are presented in Appendix C. This experiment was based on a pilot experiment (Giang & Donmez, 2015) that was used to determine the number of participants required and to refine the following experimental tasks. The methods and results from this pilot test can be found in Appendix D.

6.2.3 Experimental Tasks

In the prediction task (Figure 15), participants were presented with a visualization, were asked to predict how long it would take the rover to complete its task and responded by dragging a selection bar with their selection restricted to half minute intervals. This experimental task was similar to the workflow proposed for the STP tool in Chapter 4. After providing their prediction, participants were also asked to rate their confidence in their choice on a scale between 1 and 100.

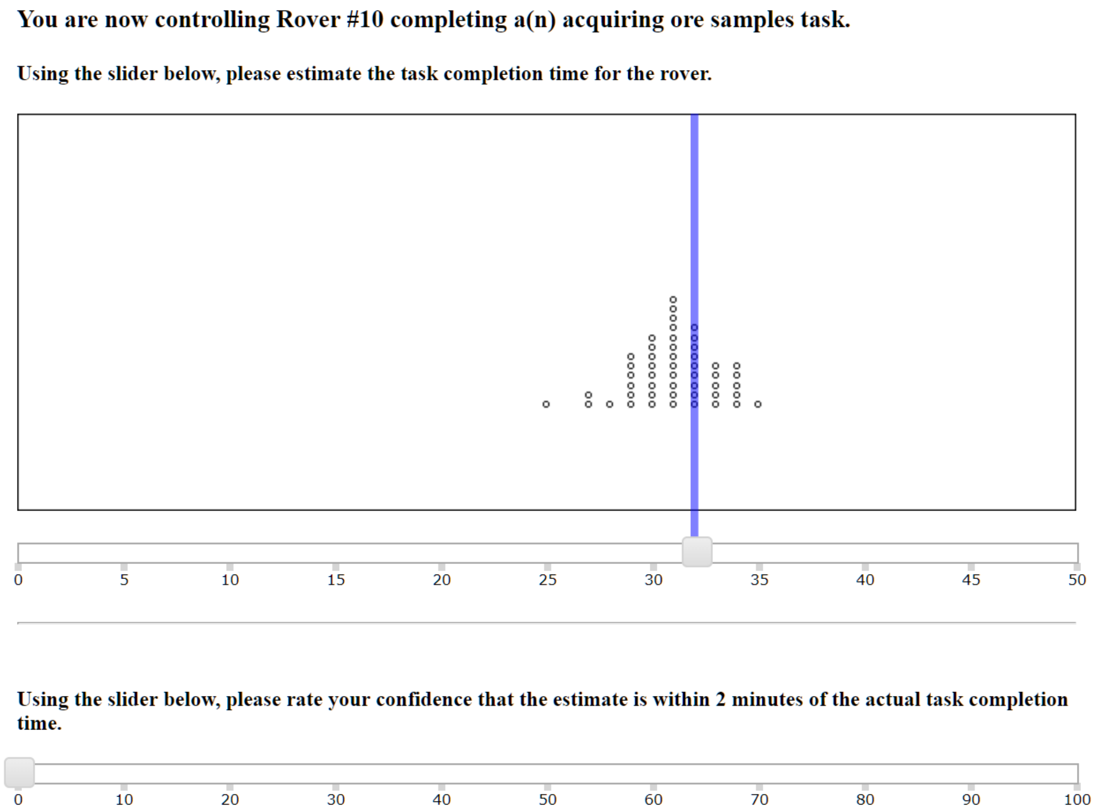


Figure 15: The prediction task

In the probability rating task (Figure 16), participants were shown their predicted task duration and were asked to estimate the probability that the actual rover task duration would be equal to or lower than their prediction. In other words, they were asked to estimate the corresponding cumulative distribution function (CDF) value, i.e., P (rover task duration ≤ predicted rover task duration). The probability rating task was designed to provide insight into the participants' internal model of the underlying probability distribution and was based on the Tak et al. (2014) study.

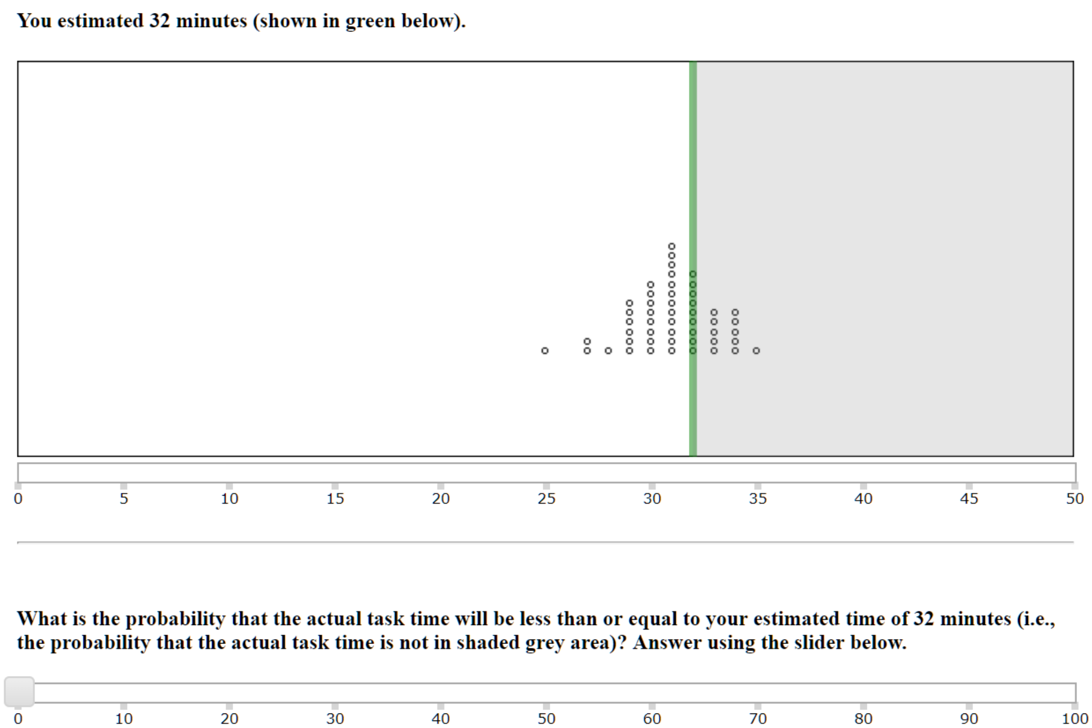


Figure 16: The probability rating task

6.2.4 Procedure

The experiment was conducted online using a custom-built website. Prior to the start of the experiment, participants were provided training on the visualizations they were assigned, which described the various graphical features of each visualization and their relation to the historical data being presented. This introduction highlighted the saliently denoted central tendency measures (Figure 14) and provided further verbal description of the central tendency and dispersion measures and the visualizations used in the experiment. Participants were then required to demonstrate their understanding of the visualizations by answering correctly to a series of test questions before being allowed to continue with the experiment. The participants then completed a practice session of two trials, one with the Median-only visualization and the other with the visualization with variability information that they were assigned. Participants were responsible for the two experimental tasks for each trial: the prediction task and the probability rating task.

Before the practice session, the participants were provided with the following instructions and interactive training about the experimental tasks: *"Your first task is to estimate how long it will* *take the rover to finish its currently assigned task, using the historical information as a guide. These estimates will be used by other mission commanders in rover logistical decisions. Therefore, it is important that your estimates are as accurate as possible. Use the slider below the visualization to select your best estimate of how long you think the rover will take to finish this rock scanning task. The blue line represents your current estimate. When you are happy with your estimate, click on the accept button. Your second task will be to estimate the probability the rover will be able to finish its current task by the estimate you provided above (i.e., how likely the rover is to finish by your estimated time). This region of interest is shown below as area that is not shaded in grey, and your estimate is shown using the green line. Use the slider below the visualization to select your probability estimate that the actual task time will be less than or equal to your estimated time (i.e., the probability that the actual task time is NOT in the shaded grey area)."* After completing the experimental trials, participants filled out a questionnaire aimed to assess their prediction strategies.

6.2.5 Dependent Variables and Statistical Analysis

Prediction behavior was assessed using four dependent variables: 1) whether participants choose the saliently presented central tendency point as their prediction, 2) the direction of the prediction relative to this central tendency point, 3) the magnitude of the deviation of the prediction form this central tendency point, and 4) their confidence in their prediction. Prediction behavior was further examined through the participants' ratings of the probability associated with their prediction, and whether the predicted values chosen by participants represented the midpoint of their internal probability model. Prediction strategies were investigated through questionnaire responses.

Statistical models were built using the SAS MIXED procedure for mixed effects linear models and GENMOD procedure for logistic and ordered logit models. Participant was used as a random factor in mixed effects models, with a compound symmetry variance-covariance structure. Generalized Estimating Equations were used in logit models to account for repeated measures. All statistical models were built using the three-way interaction between visualization condition (3 levels: Mean&SD, Boxplot, and Dotplot), variability information (Median-only vs. the other visualizations), and magnitude of variability in the dataset (small vs. large) unless otherwise stated. Apriori planned contrasts and estimated means were calculated using SAS's estimate command. Within the models reported in the results section, the visualization and variability information interaction represent the three visualizations with variability information, as well as three separate Median-only conditions for participants that experienced the Mean&SD, Boxplot, and Dotplot visualizations. The average of the three Median-only conditions was used for contrasts between the four possible types of visualizations (Median-only, Mean&SD, Boxplot, and Dotplot). Planned contrasts were also conducted between the small and large standard deviation conditions for each of the four visualization types (Mean&SD, Boxplot, Dotplot, and the average of the Median-only visualizations), and between the two levels of standard deviation for each visualization. The analysis code and SAS outputs can be found in Appendix E.

6.2.6 Hypotheses

It was hypothesized that most participants would choose the saliently denoted central tendency point in the visualization as their predicted value. It also hypothesized that regardless of the location of their prediction (i.e., on the central tendency point or not), the participants would be choosing the mid-point of their internal probability model as their prediction. Furthermore, it was hypothesized that increases in the amount of variability information provided in the visualization (based on display format) would not change participants' prediction behavior but would change their confidence in their predictions as they would have more information about the process.

6.3 Results

6.3.1 Predictions on the Salient Central Tendency Point

Overall, only 46% of predictions made with the Median-only visualization, 36% with the Mean&SD visualization, 28% with the Boxplot, and 22% with the Dotplot were on the saliently presented central tendency point (the median, mean, median, and mode, respectively). A logistic regression model was fit to examine the likelihood that the prediction was on the salient central tendency point (Figure 17). For smaller SD trials, participants were more likely to choose the salient central tendency as their prediction in the Median-only visualization than in the Boxplot (Odd Ratio – OR: 2.04, 95% Confidence Interval – CI: 1.02, 4.07) and Dotplot (OR: 3.49, 95% CI: 1.68, 7.28) visualizations. Similar results were observed for larger SD trials, but with marginal statistical significance: Median-only vs. Boxplot, OR: 2.23, 95% CI: 1.00, 5.00; Median-only vs. Dotplot, OR: 2.44, 95% CI: 0.99, 6.05. The main effect of variability magnitude was not significant, $\chi^2(1)=2.26$, p=.13. These results suggest that trials with visualizations that had the most variability information (i.e., the Boxplot and Dotplot) resulted in a lower likelihood of the salient central tendency point being chosen as the predicted rover task duration. The magnitude of variability did not appear to influence the likelihood of predicting on the salient central tendency point.

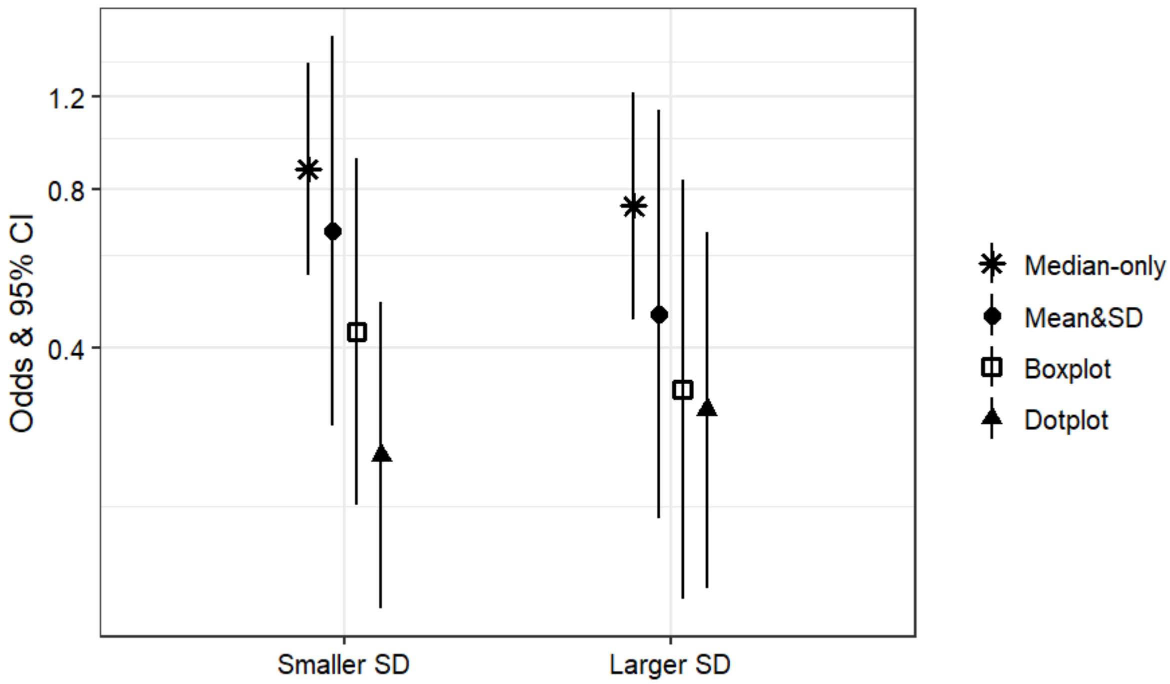


Figure 17: Statistical model estimates, odds and 95% CI, for the chosen predictions being on the saliently denoted central tendency point. Higher odds indicate greater likelihood.

6.3.2 Direction of Prediction relative to the Salient Central Tendency Point

For trials where the prediction deviated from the salient central tendency point, there were more predictions above this point than below (Figure 18). For these predictions, a logistic regression model was fit to examine the likelihood of the prediction being above the salient central tendency point as opposed to below. For smaller SD trials, participants were more likely to predict above with the Dotplot visualization than with the Median-only (OR: 14.34, 95% CI: 1.83, 112.22), Mean&SD (OR: 23.50, 95% CI: 2.66, 207.56), and Boxplot (OR: 23.50, 95% CI: 2.98, 185.31) visualizations. For larger SD datasets, participants were more likely to predict above with the Mean&SD visualization than the Dotplot visualization (OR: 2.59, 95% CI: 1.07, 6.29). Overall regardless of experimental condition, participants who deviated from the salient central tendency point in their prediction favored task durations that were longer than this value. Further, participants using the Dotplot visualization appeared to be highly influenced by the dataset that was being presented when compared to the other visualizations tested, suggesting that the additional shape information may have been guiding their predictions.

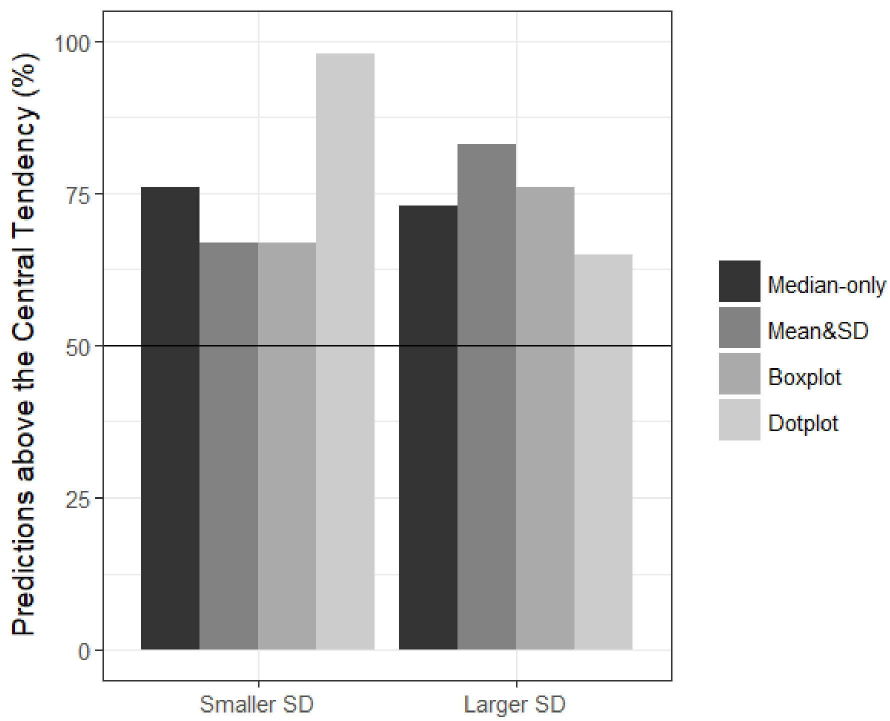


Figure 18: Percentage of trials with predictions above the salient central tendency point as opposed to below

6.3.3 Distance between Prediction and the Salient Central Tendency Point

For predictions that were not on the salient central tendency point, the average deviation from this point was 1.8 minutes for the Median-only, 1.8 for the Mean&SD, 2.1 for the Boxplot, and 1.7 for the Dotplot visualization. As mentioned earlier, participant inputs were restricted to half-minute intervals; given that these average deviations are at least three times this restriction, the deviations from the salient central tendency were not likely due to input errors.

The distance between the prediction and the salient central tendency point was divided into 4 levels with roughly equal numbers of observations: between 0.5 and 1 (n=142), between 1 and 2 (n=119), between 2 and 3.5 (n=82), and greater than 3.5 (n=114) minutes, and an ordered logistic regression model was fit to this data; the binning was performed because the data was highly non-normal. Participants tended to predict further from the salient central tendency point in larger SD trials compared to smaller SD trials for the Mean&SD (OR: 4.8, 95% CI: 2.1, 10.7), Boxplot (OR: 4.6, 95% CI: 2.6, 8.0), and Dotplot (OR: 2.7, 95% CI: 1.9, 3.7) visualizations, but not for the Median-only visualization. Thus, participants, who deviated from the salient central tendency, appeared to deviate further when variability magnitude was larger and the visualization provided variability information.

6.3.4 Confidence in Predictions

A linear mixed model was fit to participants' ratings of confidence in their predictions (Figure 19).

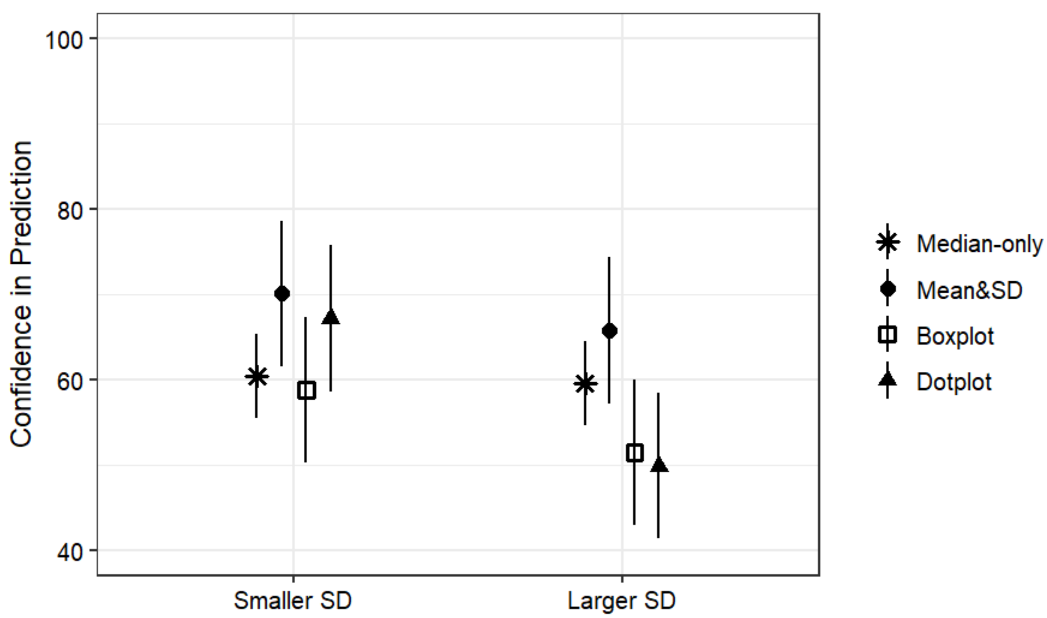


Figure 19: Statistical model estimates (means and 95% CI) for participants' ratings of confidence in their predictions

Participants were more confident in their predictions during smaller SD trials than larger SD trials for visualizations with variability information (Boxplot: Δ=7.3, 95% CI: 2.5, 12.2, Dotplot: Δ=17.3, 95% CI: 12.4, 22.2); this effect was only marginally significant for the Mean&SD visualization (Δ=4.3, 95% CI: -0.5, 9.2). The Dotplot, in particular, led to a large decrease in participants' confidence for large variability trials. For smaller SD trials, participants were less confident with the Median-only visualization than the Mean&SD (Δ= -9.7, 95% CI: -17.2, -2.2) and the Dotplot (Δ= -6.8, 95% CI: -14.3, 0.7) visualizations, with the latter comparison being only marginally significant (p=.08). For larger SD trials, participants were more confident with both the Median-only and Mean&SD visualizations than the Boxplot and Dotplot visualizations (Median-only vs. Boxplot: Δ=8.1, 95% CI: 0.6, 15.6; Median-only vs. Dotplot: Δ=9.7, 95% CI: 2.2, 17.2; Mean&SD vs. Boxplot: Δ=14.3, 95% CI: 2.2, 26.4; Mean&SD vs. Dotplot: Δ=15.9, 95% CI: 3.8, 28.0).

6.3.5 Prediction Probability

Participants were asked to estimate the probability that the actual rover task duration would be less than or equal to their own prediction. Two linear mixed models were fit to these probabilities for (1) when predictions were on the salient central tendency and (2) when they were not. When predictions were on the salient central tendency point, the corresponding probabilities were not significantly different than 50% (Figure 20, left). For the most part, when participants chose the salient central tendency point, they felt they were choosing a prediction that was the midpoint of the historical distribution. In contrast, when predictions were not on the salient central tendency point, the corresponding probabilities were significantly larger than 50% (Figure 20, right), suggesting that these participants did not consider their choice to be a midpoint of the historical distribution.

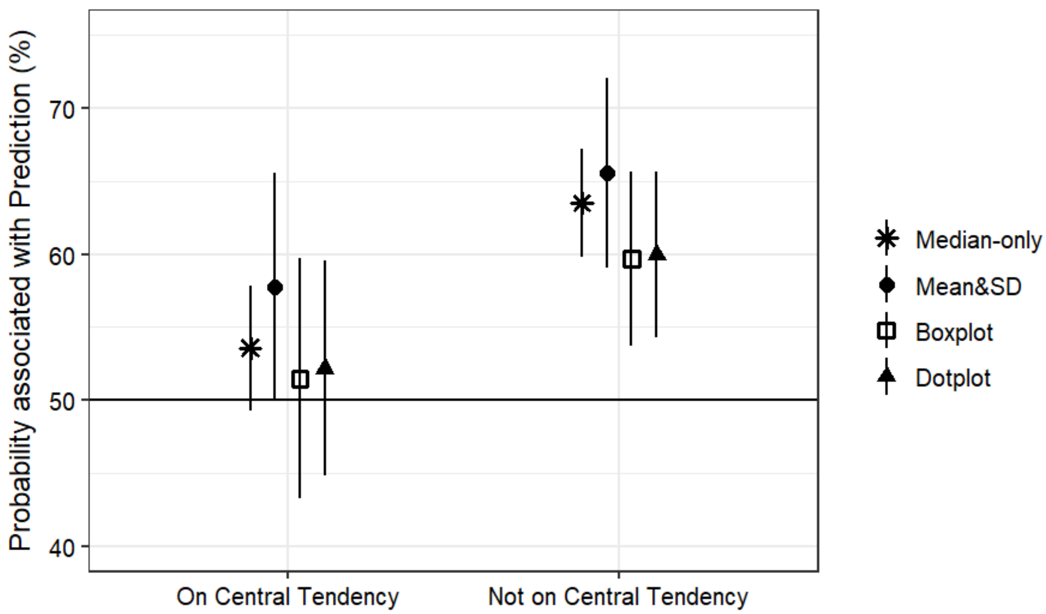


Figure 20: Statistical model estimates (means and 95% CI) for participants' probability ratings of P (rover task duration ≤ predicted rover task duration)

6.3.6 Strategies for Prediction

Participants' were asked to rate how consistent their prediction strategies were on a 5-point scale between strongly disagree and strongly agree. The majority of the participants responded with agree or strongly agree (Median-only: n=47 out of 60; Mean&SD: n=13 out of 20; Boxplot: n=15 out of 20; Dotplot: n=15 out of 20). Participants reported that they tended to rely on the mean (n=16 out of 20) for the Mean&SD visualization, and the median (n=14 out of 20) for the Boxplot visualization. However, the Dotplot visualization had more variance in terms of self-reported usage of central tendency, with nine participants stating that they used the mean, seven participants stating that they used the mode, three stating that they used the median, and the remaining person using a combination of the median and mode. Participants who stated that they always chose a central tendency point as their prediction (Median-only: n=26 out of 60; Mean&SD: n=6 out of 20; Boxplot: n=4 out of 20; Dotplot: n=7 out of 20) were found to actually do so during the experiment suggesting that there was a specific group of participants who decided to always choose the central tendency rather than switching strategies on a per-trial basis. There were overall 10 participants (out of 60 total) who always chose the saliently denoted central tendency point across both visualizations they experienced.

6.4 Discussion

Historical data are often used to help decision-makers anticipate and predict future values of important decision variables, but how individuals make these predictions is not well understood. According to the limited previous research (e.g., Ibrekk & Morgan, 1987; Nadav-Greenberg & Joslyn, 2009), individuals tend to pick the saliently presented central-tendency points as their prediction when they are presented with graphical visualizations. Similar results were expected in this experiment, given that the participants were asked to be as accurate as possible in their predictions and they were not explicitly presented any potential cost/reward structure that could bias them towards over or underestimation. The findings, however, contradict these earlier studies with the majority of the participants deviating from the saliently denoted central tendency point even when it was the only information provided on the visualization (i.e., Median-only). Furthermore, deviations from the salient central tendency point were more likely when the amount of variability information presented increased, providing evidence that the display format of the visualization influenced prediction behavior.

As a novel contribution, this study also investigated how participants' internal probability models were related to their choice of predicted value. When participants chose the salient central tendency point, they thought that this prediction was the midpoint of the historical distribution, which corresponded to our hypotheses. In contrast, when predictions were not on the salient central tendency point, the participants did not consider their prediction to be the mid-point, suggesting some other strategy was used. The experiment purposefully used an artificial scenario with little contextual information so that the participants would rely solely on the presented visualizations. The participants who chose predictions that were not on the salient central tendency point, tended to pick values above this point suggesting a potential risk-averse prediction strategy.

Since participants were asked to produce a prediction that would be communicated to and used by other team members for their scheduling tasks, some participants might have embedded their own biases into the task, despite the fact that they were asked to predict as accurately as possible. Previous research has found that individuals adjust their predictions in response to varying cost and reward structures, although not in an optimal manner (Mannes & Moore, 2013). As for time predictions, Burt and Kemp (1994) found that individuals tended to overestimate how long it would take them to complete everyday tasks because overestimations help avoid the stress associated with not being able to complete the task by the predicted time. It is important to note that predictions can also be biased to be more risky as is often found in situations such as the planning fallacy (Kahneman & Tversky, 1982a), and in predictions of costs for public projects (Flyvbjerg et al., 2002), where optimistic predictions may provide the greatest immediate value for the decision maker. In the current study, participants were asked to make predictions about rover task durations that would be used to support task scheduling, thus a more pessimistic prediction may have been favored.

The hypothesis of risk-averse predictions by those who deviated from the salient central tendency point is supported by their probability ratings. When asked the CDF value corresponding to their prediction, these participants indicated values greater than 50%. Thus, the participants were likely overestimating such that they considered their prediction to have a higher than 50% chance of being longer than the actual task duration. Instead of focusing on producing an “accurate” prediction (i.e., choosing a prediction to reduce the long-term error), participants may have confounded the prediction task with the possible consequences that may arise from the use of the prediction, even when the current experiment stressed the accuracy of predictions.

It is also possible that the relatively simple structure of the task biased the participants towards adopting strategies that made use of all the information that they were presented rather than simply picking the salient central-tendency point. For example, with the Boxplot and Dotplot visualizations, the participants could have attempted to estimate the mean which was not saliently denoted. In fact, the post-experiment provided some evidence for this potential strategy, in particular for the Dotplot visualization. For the Median-only condition, the participants could have used the range of the x-axis (kept constant throughout the experiment) as an indicator of potential process variability. It is also possible that dispositional factors, such as numeracy ability or risk tolerance, of these participants resulted in the differing strategies. While participants were screened for their knowledge of statistics, there is likely large variability in how participants applied this knowledge to their prediction behavior. Future research should examine how these individual factors may impact prediction strategies in the presence of uncertainty information. It is also important to note that the study was conducted online, resulting in a less controlled environment for the participants. However, the use of the online platform also allowed for a more diverse participant sample than those normally used in experimental studies (e.g., college students).

One advantage of the visualizations that provided variability information was that participants were able to calibrate their confidence in their predictions based on the magnitude of the variability in the historical data; when there was more variability, the confidence ratings were lower. When confronted with uncertainty, individuals adopt a number of different strategies, including ignoring uncertainty, planning for worst-case scenarios, and gathering more information (Lipshitz & Strauss, 1997), and differences in confidence may change the strategy used to cope with uncertainty.

Surprisingly, visualizations with variability information did not necessarily lead to higher confidence ratings compared to the Median-only visualization. This unexpected finding may be due to the Median-only visualization being always presented first, and participants readjusting their confidence when they could actually see the historical variability in the later trials. Furthermore, the Mean&SD visualization resulted in the highest confidence ratings of all the visualizations, which may be due to the bounds indicated by standard deviation appearing much shorter than those indicated by the range of historical data depicted in Boxplot and Dotplot visualizations, especially given that the range of the x-axis was kept constant throughout the experiment. In addition, previous research has shown that error bars are often misinterpreted (Correll & Gleicher, 2014); gradient plots (Correll & Gleicher, 2014) that make use of opacity as a visual indicator of uncertainty and other visualization methods, such as hypothetical outcome plots (Hullman, Resnick, & Adar, 2015), have been suggested as alternative methods to represent uncertainty. In addition, the four visualizations used in this study represent commonly-used methods for presenting historical data, however, the particular ways (i.e., the thickness of the standard deviation bars or the height of the Boxplot) they were represented in the experiment may have impacted prediction behavior. Thus, further research is needed to examine other design choices for the visualizations studied in this experiment and other types of uncertainty visualizations.

6.5 Conclusion

It was hypothesized that most participants would choose the saliently denoted central tendency point in the visualization as their predicted value and that regardless of the location of their prediction (i.e., on the central tendency point or not), the participants would be choosing the mid-point of their internal probability model as their prediction. When some participants chose the saliently denoted central tendency point as their prediction, it corresponded to the mid-point of their internal probability models. However, the majority of participants deviated from the central-tendency point and when they did they appeared to be choosing predictions that were more conservative than the mid-point.

It was also hypothesized that increases in the amount of variability information provided in the visualization (based on display format) would not change participants' prediction behavior but would change their confidence in their predictions. The increase in variability information did change prediction confidence as expected but it also changed prediction behavior.

In conclusion, the major findings of this chapter were that the display format influences prediction behavior; more variability information leads to more deviations from the saliently presented central tendency point. However, even with no variability information, people do deviate from the point-estimate; e.g., by making conservative predictions suggesting that prediction behavior is also tied to factors outside of the information presented in the visualization. The presence of variability information, especially in the Boxplot and Dotplot visualizations, allowed participants to adjust their confidence in their predictions, which may have beneficial effects on subsequent decisions made using the predictions.

The results of this study have implications for designers of decision support systems that support forecasting and predictions using historical data. Within the context of medical dispatch, the prediction of patient transfer times for different transport methods (i.e., air vs. ground transfer) is an important decision variable. In the past, these time predictions have largely relied on dispatcher experience and intuition, and the introduction of decision-aids to support this aspect of the dispatch task can help standardize and improve prediction outcomes. Providing variability information in addition to point-estimates can be useful when users have relevant-contextual information and can utilize this information to improve the prediction that is provided by a predictive-aid (e.g., inclement weather slowing patient transfer times). However, it appears that users are more likely to factor in contextual information that can also degrade prediction accuracy when they are presented with more variability information. The influence of contextual information on prediction behavior with the use of historical data visualizations is examined with a second experimented presented in the following chapter.

Chapter 7

7 Influence of Contextual Information on Time Prediction

Field studies reported in Chapter 3 revealed that dispatch decisions are influenced by many contextual factors and two specific categories of contextual information that may influence transfer time predictions were identified. Consequence-information refers to factors that do not change the underlying process of the patient transfer but are associated with the cost or rewards for certain outcomes (e.g., if the transfer takes longer than 45 minutes then the crew will need to be paid overtime). Likelihood-information refers to factors that should change the underlying transfer process leading to different transfer times (e.g., snow results in slower than normal transfer times). As discussed in Chapter 5, the identified types of contextual information may play a role in how users make predictions of time using visualizations of historical data.

Although the first experiment reported in Chapter 6 aimed to remove contextual information as much as possible, participants still appeared to have incorporated contextual information that changed their prediction strategies for rover time prediction; participants' time prediction behavior was not solely dependent on finding the central tendency of the visualized data or the mid-point of their internal probability models. This chapter presents an exploratory study that examines whether and how the likelihood- and consequence-information changes time prediction behavior and strategies when individuals are provided with historical data visualizations. Three context information conditions were examined: No contextual information (no-context), consequence-information, and likelihood-information. Given that the focus of this experiment was context, a medical dispatch scenario was adopted in place of rover time prediction. The contribution of this chapter is preliminary evidence that prediction strategies differ based on the type of contextual information provided.

7.1 Experimental Scenario and Visualizations Tested

In contrast to the experiment in Chapter 6, the study presented in this chapter used the medical dispatch domain where participants were asked to take on the role of a medical dispatcher responsible for predicting patient transfer times. Participants were asked to predict the duration of a patient transfer using visualizations of historical patient transfer times. The experiment in

Chapter 6 showed that even in an unfamiliar scenario, such as rover time prediction, participants appeared to bring their own biases (i.e., risk-averse prediction behavior) to the experimental task, thus the patient transfer scenario was selected to increase the validity of the findings of this study in supporting the design of support tools for Ornge. Furthermore, the use of a patient transfer scenario allowed for the utilization of contextual information types reported in Chapter 3 that were identified from observations with Ornge medical dispatchers.

The two of the four visualization conditions tested in Chapter 6 were adopted for the current study in order to reduce experimental complexity: a baseline Median-only visualization that had no uncertainty information, and a Boxplot visualization that showed variability information. The results of Chapter 6 showed that visualizations with uncertainty information, such as the Boxplot visualization, resulted in different prediction behavior than predictions made with only a point-estimate (i.e., the Median-only visualization). The Boxplot visualization was selected because it provided both variability and skewness information and used the same measure of central tendency (i.e., the median) as the Median-only visualization.

7.2 Methods

7.2.1 Experiment Design

The experiment used a 2x3x2 factorial design with visualization type (Median-only vs. Boxplot), context information (no-context, consequence-information, and likelihood-information), and the variability magnitude of the dataset (smaller vs. larger SD) as within-subjects factors (Table 3). Each participant was presented with six trials with the Median-only visualization first as a baseline before completing six trials with the Boxplot visualization, resulting in 12 trials per participant. Within each visualization condition, six trials corresponded to the six different combinations of the variability magnitude and context information conditions (3x2). The order of presentation of these six trials within each visualization condition was randomized.

Table 3: Experimental design for contextual information experiment

<table>
<tr><th colspan="3">Block 1: Trials 1-6</th><th colspan="3">Block 2: Trials 7-12</th></tr>
<tr><td rowspan="6">Median-only</td><td rowspan="3">Smaller SD
Datasets 1-3</td><td>No Context</td><td rowspan="6">Boxplot</td><td rowspan="3">Smaller SD
Datasets 1-3</td><td>No Context</td></tr>
<tr><td>Consequence</td><td>Consequence</td></tr>
<tr><td>Likelihood</td><td>Likelihood</td></tr>
<tr><td rowspan="3">Larger SD
Datasets 4-6</td><td>No Context</td><td rowspan="3">Larger SD
Datasets 4-6</td><td>No Context</td></tr>
<tr><td>Consequence</td><td>Consequence</td></tr>
<tr><td>Likelihood</td><td>Likelihood</td></tr>
</table>

The participants were presented with a Median-only visualization or a Boxplot of patient transfer times, and were told that a given visualization was generated from relevant historical patient transfers. Overall, six unique datasets were created as the historical data for the six trials experienced by participants; these six datasets were repeated across the two visualization conditions. Half of the datasets had smaller SD and half had larger SD (Appendix F). The datasets were created from normal distributions, each consisting of 50 observations that were evenly spaced throughout the distribution (e.g., 1st, 3rd, 5th... 99th percentiles) to ensure that the datasets appeared normal and symmetric. Each observation was rounded to the nearest half minute. Smaller SD datasets came from distributions with SD of 2 minutes, the larger SD datasets came from distributions with SD of 5 minutes.

Contextual information was manipulated by presenting a short excerpt above the visualization (Figure 21) and represented information that the user of the decision-aid would have about the situation that is not explicitly encoded in the system (i.e., the visualization). The excerpts used for each condition can be found in Table 4. The consequence-information described the situation (e.g., end of shift and patient condition), and a time threshold that demarked a change in the outcome of the transfer (e.g., 26 or 30 minutes). The thresholds were always one SD above the median of the historical data. The likelihood-information consisted of a description of the current situation that may impact the underlying process of the patient transfer (e.g., inexperienced team or traffic), along with an adjustment magnitude (e.g., 2 or 5 minutes slower than normal). These adjustment magnitudes were set to one SD of the underlying dataset and were always slower than the median (i.e., above the median). Thus, both types of context information referred to the same time value (i.e., one SD above the median), but framed the information in different ways. Two different stories were created for each context condition to reduce the similarity between trials within a given visualization type; these two stories were however repeated across the two visualizations. How the stories were paired with the experimental conditions was counterbalanced as much as possible.

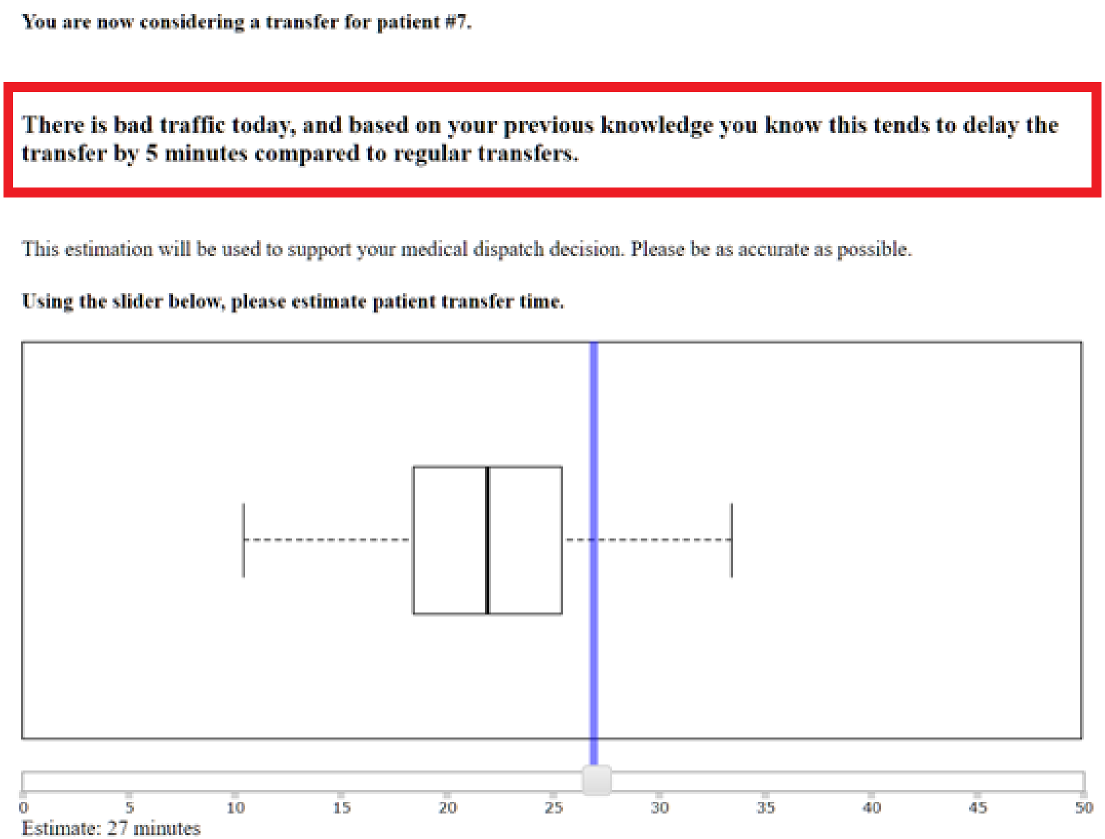


Figure 21: Example of the prediction task with the Boxplot visualization, with the likelihood-information presented above the visualization of the historical data (highlighted with a red box for this figure).

Table 4: The story excerpts used for each context condition

Context	No-Context	Consequence	Likelihood
Story 1	*"No additional information is known about the transfer."*	*"The paramedic crew is at the end of the shift, if the transfer takes longer than 26 minutes then they will reach the end of the shift and will need to be replaced by a new crew or paid overtime."*	*"The dispatch team is new and inexperienced, and based on your previous knowledge you know that they tend to complete this transfer 2 minutes slower than most crews."*
Story 2	*"No additional information is known about the transfer."*	*"Based on the medical status of the patient, if the transfer takes longer than 30 minutes then there is a higher chance of loss of life or limb."*	*"There is bad traffic today and based on your previous knowledge you know this tends to delay the transfer by 5 minutes compared to regular transfers."*

7.2.2 Participants

Participants were recruited similarly to the first experiment, using a paid screening questionnaire on Amazon MTurk, which was open to all MTurk Workers in the United States and Canada with Masters Qualification. Participants were screened for having normal or corrected-to-normal vision and normal color perception through self-reports. In addition, the screening questionnaire provided definitions and examples of central tendency (i.e., mean, median, and mode) and dispersion (i.e., range, interquartile range, and standard deviation) concepts. Participants were tested on these concepts and had to score at least 75% to qualify for the study. Overall, 150 participants answered this initial screening questionnaire, of which 123 qualified and were invited to participate in the study. The screening questionnaire took approximately 15 minutes and participants were compensated US$1.5 for filling out the screening questionnaire.

Eighty participants completed the study, but the data from one participant was removed due to missing responses for one of the experimental tasks. The remaining 79 participants' data were used for analysis (41 male, 38 female, mean age = 38.9 years, SD = 11.0). Participants scored an average of 92% (SD = 7%) on the screening questionnaire. Participants came from a variety of educational backgrounds ranging from high school to graduate school, with the majority of participants (n = 56) having completed post-secondary education. 47 participants had taken at least one probability or statistics course at the post-secondary level and a further 12 had completed a probability or statistics course at the high school level. The experiment took 45 minutes and participants were compensated US$8.5. The study was approved by the University of Toronto Research Ethics Board. Informed consent was obtained from each participant for both the screening questionnaire and the actual experiment. Relevant experimental materials can be found in Appendix F.

7.2.3 Experimental Tasks

Similar to the first experiment, participants were given a prediction task and a probability rating task. In the prediction task (Figure 21), participants were presented with a visualization, were asked to predict how long it would take for the patient transfer, and responded by dragging a selection bar with their selection restricted to half minute intervals. After providing their prediction, participants were also asked to rate their confidence in their choice on a scale between 1 and 100. In the probability rating task (Figure 22), participants were shown their predicted transfer time and were asked to estimate the probability that the actual patient transfer would be equal to or lower than their prediction. In other words, they were asked to estimate the corresponding cumulative distribution function (CDF) value, i.e., P (patient transfer time ≤ predicted patient transfer time).

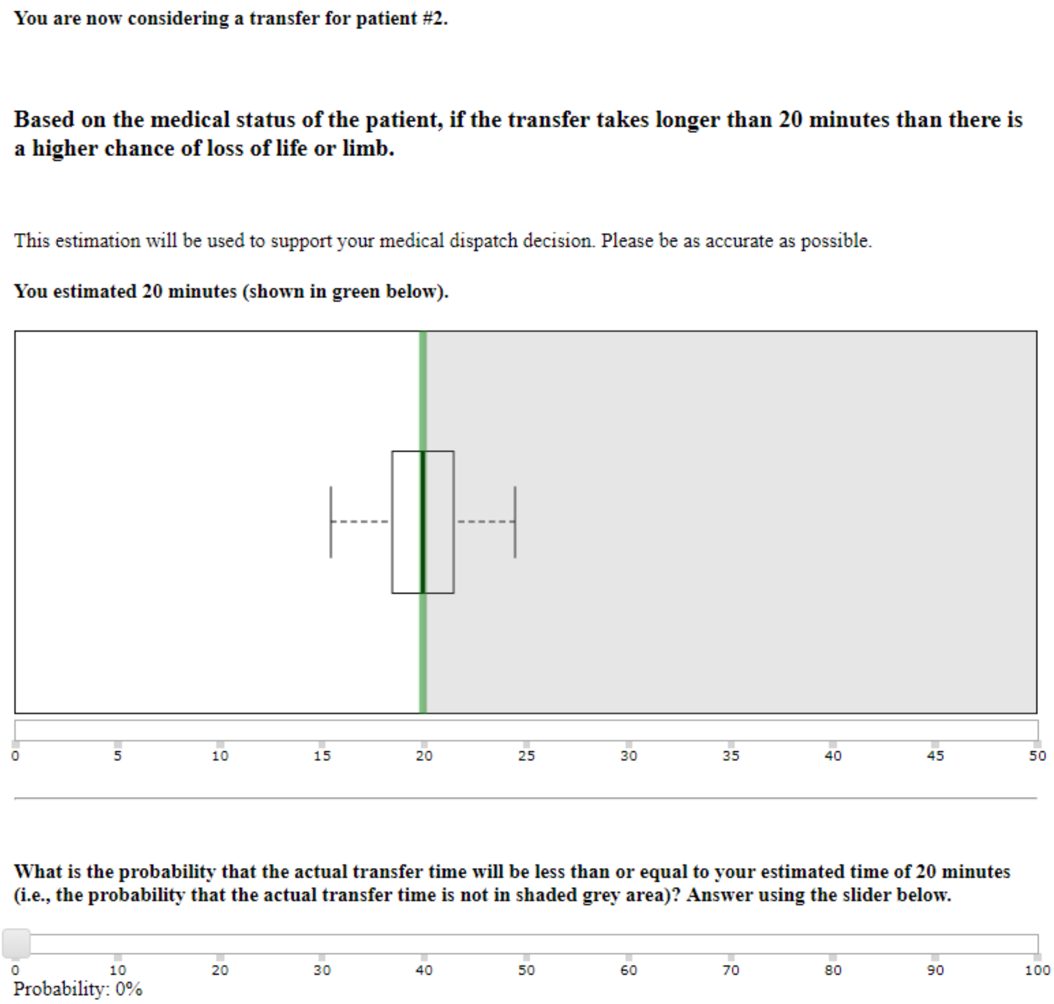


Figure 22: Example of the probability rating task with the Boxplot visualization and consequence-information.

7.2.4 Procedure

Prior to the start of the experimental trials, participants were presented with the following introduction to the medical dispatch scenario.

> *"You are taking the role of a medical dispatcher. Patients often require transfers between hospitals, and medical dispatchers organize these transfers by dispatching the best vehicle for completing the transfer. Estimates of the time required to transfer a patient are important to medical dispatchers. Estimates are used in comparing different transfer options (such as using a helicopter or*

> *land ambulance) and are also provided to hospitals so they can prepare for the patient. This experiment focuses on how you generate time estimates.*
>
> *During the experiment, you will be asked to support a series of medical dispatch decisions by estimating how long you think it will take to transfer the patient.* ***To help you with these estimates, you will be presented with historical data about how long it has taken to transfer patients in the past, along with information about the specific patient transfer that you are considering.*** *Patient transfer times are dependent on many factors, and vary even for transfers between the same pair of hospitals. For example, traffic conditions, the crew assigned to the transfer, and the weather can impact the transfer time. For each patient, you will be presented with a graph that summarizes all of the historical transfer time data for transfers between the hospitals being considered…*
>
> *…***On some occasions, you may also be provided additional information about the current patient transfer.*** *For example, you may be told information about specific patient relevant deadlines or targets, or information about the current situation (e.g., weather or traffic conditions). It is important to remember that this information only applies to the patient transfer being considered, while the historical data is a summary of many different transfers for different patients."*

Participants were also given training with the Median-only and Boxplot visualizations at the beginning of the experiment. The training described the various graphical features of each visualization and their relation to the historical data being presented. This introduction highlighted the saliently denoted central tendency measures (Figure 14) and provided further verbal descriptions of these measures. Participants were then required to demonstrate their understanding of the visualizations before continuing with the experiment. The participants then completed a practice session of two trials, one with the Median-only visualization with likelihood-information and the other with the Boxplot with consequence-information. After completing the experimental trials, participants filled out a questionnaire to help assess their prediction strategies. The questionnaire asked about predictions made under the Median-only and Boxplot visualizations in separate sections. First, participants were asked how much of an impact each type of context information had on their predictions, rated on a 5-point scale between Strongly Disagree and Strongly Agree. Participants were then asked whether their predictions were an optimistic, average, or pessimistic case for each context condition.

7.2.5 Dependent Variables and Statistical Analysis

The prediction behavior was assessed using the same dependent variables as Chapter 6: 1) whether participants choose the saliently presented central tendency point as their prediction, 2) the direction of the prediction relative to this central tendency point, 3) the magnitude of the deviation of the prediction from this central tendency point, and 4) their confidence in their prediction.

The analysis of the probability rating task was conducted differently than it was done for the first experiment. Within each of the three context conditions, the participants were categorized into different strategy groups (e.g., median picked as the prediction along with a 50% CDF value) based on the location of their prediction relative to the saliently presented central tendency point (i.e., median) and their probability rating. This categorization helped with the assessment of whether and how participants' strategies changed with context information. Participants' observed strategies were also analyzed with respect to what they indicated in their questionnaire responses (i.e., average, pessimistic, optimistic).

Statistical models were built using the SAS MIXED procedure for mixed effects linear models and GENMOD procedure for logistic and ordered logit models. Unless otherwise specified, statistical models were built with contextual information, visualization type, magnitude of variability of the dataset, and their interactions as fixed factors. Participant, nested under the visualization type, variability magnitude, and context interaction, was used as a random factor in mixed effects models, with a compound symmetry variance-covariance structure. Generalized Estimating Equations were used in logit models to account for repeated measures. The significance of factors was tested using Type III Wald chi-square tests and models were selected using backwards selection with insignificant factors being dropped from subsequent models. Contrasts and estimated means were calculated using SAS's estimate command. The analysis code and SAS outputs can be found in Appendix G.

7.2.6 Hypotheses

It was hypothesized that the contextual information changes participants' prediction behaviors. When no-information is provided, participants would be more likely to rely on the presented data and would choose the central tendency of the visualization as their prediction. The first experiment showed that participants tended to overestimate rover task durations potentially because they were considering consequences of underestimation although they were not explicitly provided any such context information. Based on these results, for the current experiment, it was hypothesized that in the presence of consequence-information, participants would adjust their predictions to avoid underestimating transfer time with probability ratings greater than 50%, demonstrating a potential bias. Finally, when presented with likelihood-information, participants would center their internal models towards the suggested adjustment, resulting in changes in the predicted time but a reported probability rating that is near 50%.

7.3 Results on Prediction Behavior

7.3.1 Predictions on the Salient Central Tendency Point

Overall, 72% of predictions made with no-context, 39% of predictions made with consequence-information, and just 3% of predictions made using likelihood-information were on the saliently presented central tendency point of the visualizations (i.e., the median). A logistic regression analysis found that only the effect of context was significant, $\chi^2(2)= 134.8$, $p<.0001$. Thus, there was no evidence that the visualization type or the magnitude of the variability changed whether participants chose the saliently presented central tendency point as their prediction. Trials with no-context resulted in more predictions on the median than consequence-information (OR: 4.0, 95% CI: 2.8, 5.7) and likelihood-information (OR: 96.7, 95% CI: 38.9, 240.4), and consequence-information resulted in more predictions on the median than likelihood-information (OR: 24.2, 95% CI: 9.6, 61.2), providing evidence that the presence of context information strongly influenced whether participants chose the saliently presented central tendency point as their predictions, and the three context information conditions differed in the magnitude of their effects.

7.3.2 Direction of Predictions relative to the Salient Central Tendency Point

For trials where the predictions deviated from the salient central tendency point, 72% of the predictions were above the saliently presented central tendency point, but it appears that this was mostly due to the likelihood-information condition (Figure 23). A logistic regression analysis revealed that only the main effects of context, $\chi^2(2)= 43.4$, $p<.0001$, and variability magnitude of the dataset, $\chi^2(1)= 5.9$, $p=.02$, were significant. As expected, the likelihood-information condition, which indicated that the process is likely to be delayed due to traffic or a slow crew, resulted in more predictions above the median than both the no-context (OR: 8.0, 95% CI: 3.8, 16.8) and the consequence-information (OR: 7.7, 95% CI: 4.1, 14.3) conditions. Contrary to the expectations, there were no differences between the consequence-information and the no-context conditions.

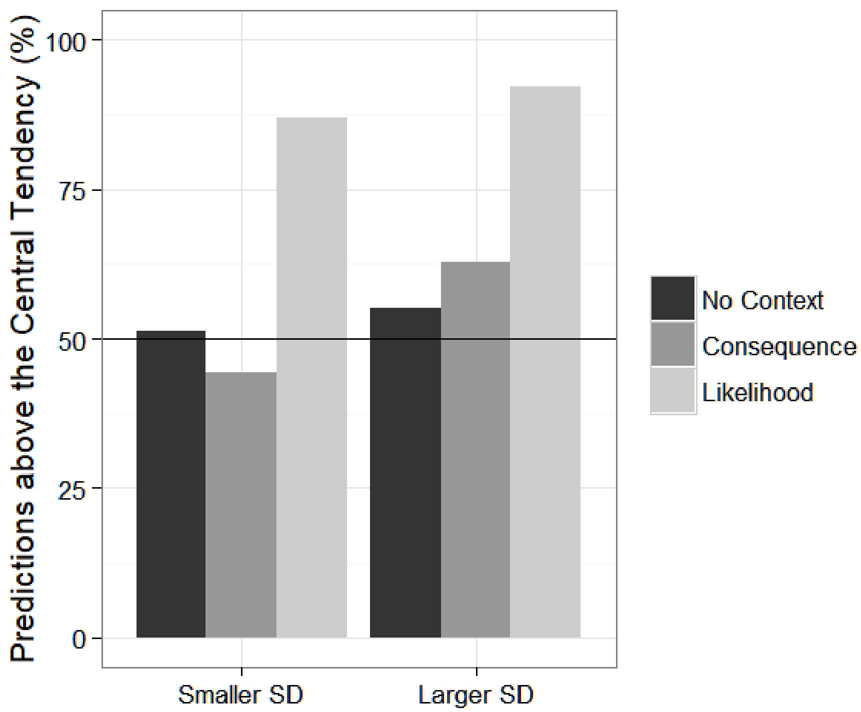


Figure 23: Percentage of trials with predictions above the salient central tendency point as opposed to below

7.3.3 Distance between Prediction and the Salient Central Tendency Point

For predictions that were not on the salient central tendency point, the distance between the prediction and the salient central tendency point was divided into 4 levels with roughly equal numbers of observations: between 0.5 and 1 (n=193), between 1 and 2 (n=170), between 2 and 4 (n=94), and greater than 4 (n=135) minutes; the binning was performed because the data were highly non-normal. Figure 24 shows the number of predictions within each category across the different variability magnitude and context conditions.

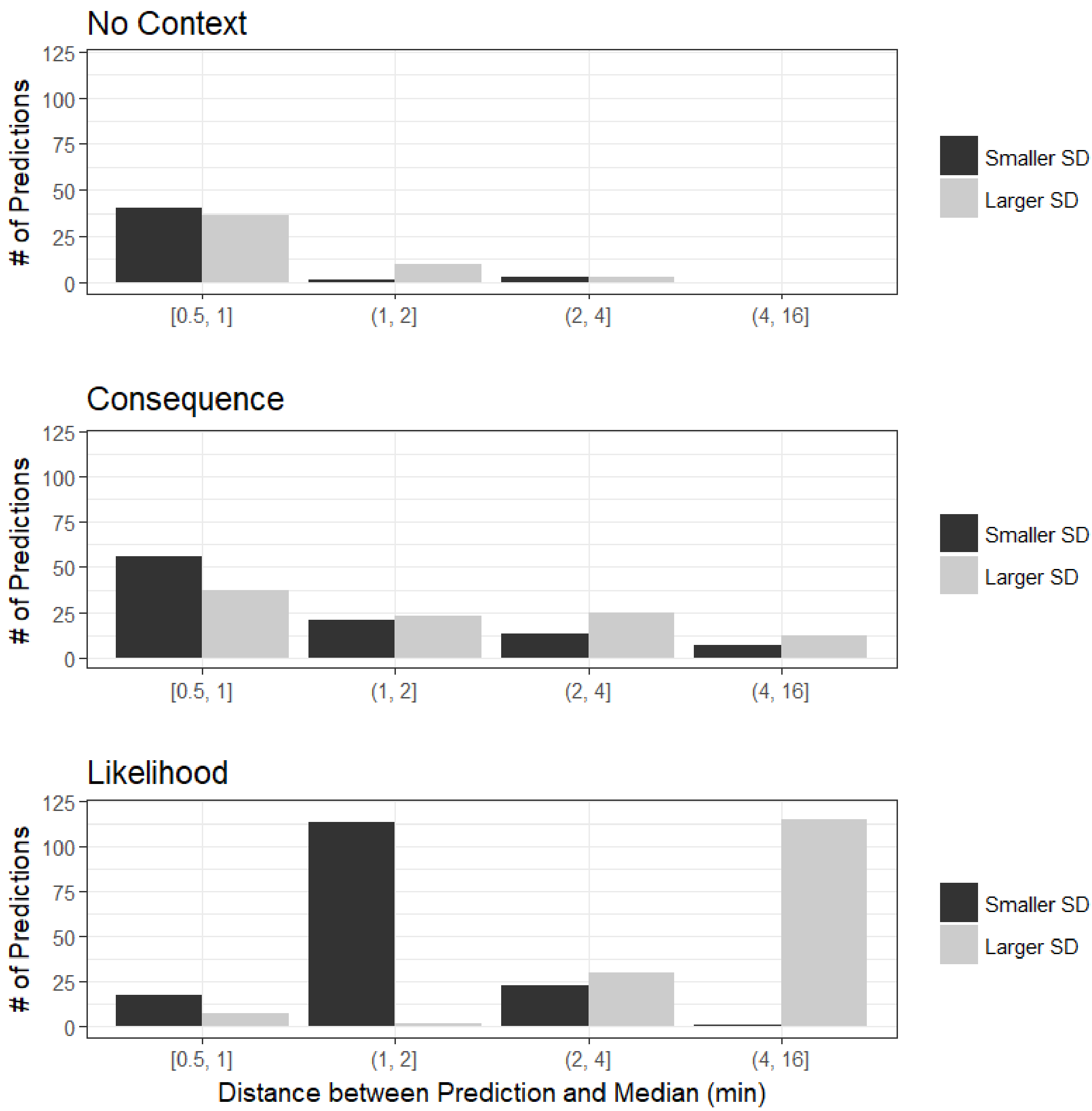


Figure 24: Number of predictions within each distance category for each level of variability magnitude and context information

The adjustments away from the central tendency were found to vary based on the context information and variability magnitude; an ordered logistic regression model found that the interaction between context and variability magnitude was significant, $\chi^2(2)=33.8$, $p<.0001$, as were the main effects of context information, $\chi^2(2)= 116.7$, $p<.0001$, and variability magnitude of the dataset, $\chi^2(1)= 43.5$, $p<.0001$. Participants adjusted their predictions based on the variability magnitude, with trials with smaller SD resulting in predictions closer to the median than trials with larger SD for the no-context (OR: 0.07, 95% CI: 0.01, 0.58), consequence-information (OR: 0.33, 95% CI: 0.17, 0.66), and likelihood-information (OR: 0.03, 95% CI: 0.02, 0.05) conditions.

Differences were also found between the three context conditions. In trials with smaller SD, no-context resulted in predictions closer to the median than consequence-information (OR: 0.03, 95% CI: 0.004, 0.21) and likelihood-information (OR: 0.01, 95% CI: 0.001, 0.062); consequence-information also resulted in predictions closer to the median than likelihood-information (OR: 0.32, 95% CI: 0.17, 0.60). Similar results were found for trials with larger SD: no-context resulted in predictions closer to the median than consequence-information (OR: 0.13, 95% CI: 0.06, 0.32) and likelihood-information (OR: 0.003, 95% CI: 0.001, 0.003), and consequence-information resulted in predictions closer to the median than likelihood-information (OR: 0.03, 95% CI: 0.01, 0.06).

7.3.4 Confidence in Predictions

Participants' confidence in their predictions appeared to differ between the visualization types and variability magnitudes. A linear mixed model fit to participants' confidences in their prediction found that the main effects of visualization type ($\chi^2(1)$= 15.01, p=.0001 and variability magnitude ($\chi^2(1)$= 12.05, p=.0005), and their interaction were significant, $\chi^2(1)$= 8.37, p=.004.

Participants reported having the lowest confidence in trials with larger SD datasets in the Boxplot condition compared to all other combinations of visualization type and variability magnitude (smaller SD-Boxplot: Δ= 5.4, 95% CI: 3.0, 7.7; smaller SD-Median-only: Δ = 6.2, 95% CI: 3.9, 8.6; larger SD-Median-only: Δ = 5.7, 95% CI: 3.4, 8.1). Thus, uncertainty information helped individuals lower their confidence in the right situations (i.e., when the magnitude of the variability in the dataset was larger). However, participants appeared to have high confidence even without the presence of uncertainty information in the Median-only visualization condition. Interestingly, context information did not affect participants' confidence in their predictions even though the context information was provided as an additional source of information about the current situation and participants' actual predictions were strongly influenced by the context information.

7.3.5 Summary of Prediction Behavior Results

Overall, it appears that the three types of contextual information had large and differing effects on participants' prediction behavior. In the no-context condition, participants were much more likely to choose the saliently presented central tendency point. Consequence-information changed the prediction behavior, with fewer participants choosing the central tendency, but with a relatively even distribution of predictions above and below the salient central tendency point. This was in contrast with the original hypothesis that consequence-information would result in predictions above the median indicating a potential bias to avoid underestimation. As in line with the hypothesis, the likelihood-information appeared to strongly influence prediction behavior away from the central tendency point towards the time adjustment suggested by contextual information. The following section further explores the reasons for observed behaviors through an analysis of participants' internal probability models.

7.4 Results on Prediction Strategy

7.4.1 Probability Rating and Prediction Location

Figure 25 provides an overview of prediction strategies by presenting the location of participants' predictions relative to the salient central tendency point (i.e., the median) and their probability ratings. Within each of the three context conditions, each participant was assigned to nine different strategy categories based on these two dimensions: location with respect to central tendency (below, on, above) x probability rating (below 50%, 50%, above 50%). For a context condition, participants were assigned into one of these nine categories which represented the majority of their predictions. The breakdown is presented in Table 5. When there was no majority, the participant was assigned to a tenth category (labeled "Other" in Table 5).

Within the nine categories, four strategy groups are of particular interest: *Strategy-1.* Participants who chose the central tendency point and a probability rating of 50% likely utilized only the visualized historical data; *Strategy-2.* Participants who chose a value larger than the central tendency point with a probability rating larger than 50% likely adopted a pessimistic prediction strategy; *Strategy-3.* Participants who chose a value smaller than the central tendency point with a probability rating less than 50% likely adopted an optimistic prediction strategy; *Strategy-4.* Participants who did not pick the central tendency point but picked a probability rating of 50%

likely adjusted their internal probability models based on the contextual information and selected what they consider to be the central tendency for the given situation. This categorization helped with the assessment of whether and how participants' strategies changed with context information. Participants' observed strategies were also analyzed with respect to what they indicated in their questionnaire responses (i.e., whether their prediction was average, pessimistic, or optimistic). Our expectation was that the participants would tend to exhibit *Strategy-2* in the consequence-information condition and *Strategy-4* in the likelihood-information condition.

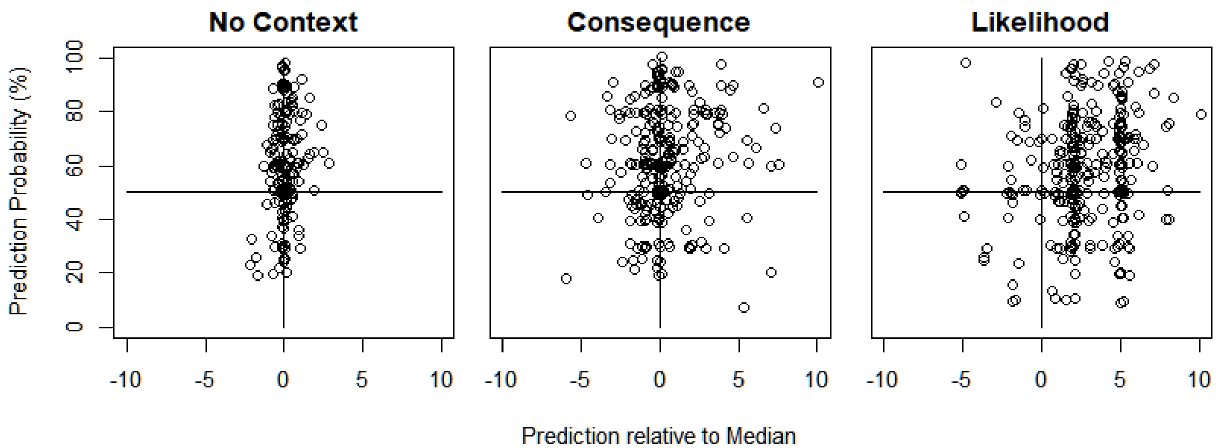


Figure 25: Predictions made by participants assessed across two dimensions: their location relative to the salient central tendency point and the prediction probability.

Table 5: Participant prediction strategies across the three context conditions. The strategy used by the plurality of participants is bolded.

		No-context Other = 27			**Consequence** Other = **45**			**Likelihood** Other = 27		
		Prediction Location								
		Below	*Median*	*Above*	*Below*	*Median*	*Above*	*Below*	*Median*	*Above*
Prediction Probability	*>50%*	5	9	7	6	8	9	0	0	**36**
	50%	0	**29**	0	0	8	0	1	0	3
	<50%	0	2	0	2	1	0	0	0	12

The results show the strong effect the three different context conditions had on prediction strategies. In the no-context condition, the most common prediction strategy used by participants was *Strategy-1*. For the consequence-information condition, the majority of the participants' prediction strategies varied between trials and thus these participants were assigned to the

"Other" category. It appears that participants changed their prediction strategies based on the consequence-information in comparison to the no-context condition, but the manner in which each participant used the information varied across the sample; it's possible that the varying stories used across trials may have introduced this variability. Finally, for likelihood-information condition, the participants tended adopt *Stategy-2* in contrast to the expectations.

After the trials, participants were asked whether their predictions were an optimistic, average, or pessimistic case for each context condition, and provided separate responses for the Median-only and Boxplot visualizations. The type of prediction (i.e., optimistic, average, or pessimistic) indicated by the majority of participants for each prediction strategy category is listed in Table 6 (the percentages used to create Table 6 can be found in Table 11 in Appendix H).

Table 6: The type of prediction indicated by the majority of participants within each strategy category: optimistic (Opt), average (Ave), or pessimistic (Pes). The type of prediction represents the majority for both visualizations unless otherwise noted. Categories with no participants are labelled with 0, while the number of participants within each other strategy category is presented in Table 5.

		No-context Other = Ave			**Consequence** Other = Opt			**Likelihood** Other = Ave		
		Prediction Location								
		Below	*Median*	*Above*	*Below*	*Median*	*Above*	*Below*	*Median*	*Above*
Prediction Probability	*>50%*	*	Ave	Pes	Opt	Ave	Pes	0	0	Pes
	50%	0	Ave	0	0	Ave	0	Ave	0	Ave
	<50%	0	Ave	0	Opt	Ave	0	0	0	Ave

* Median-only: Opt; Boxplot: Ave

It appears that prediction strategies reported by the participants differed by context. In general, if participants were using their internal probability model to select their predicted values as was hypothesized in the framework proposed in Chapter 5, one would expect that predictions that had a probability rating less than 50% to be optimistic, those that were on 50% to be an average case, and those larger than 50% to be pessimistic. However, the data did not appear to support this hypothesis; instead, the self-reported prediction type (i.e., optimistic, average, or pessimistic) appeared to be related to the location of the predicted value relative to the median rather than the probability rating assigned to it.

It was hypothesized that in the presence of consequence-information, participants would adjust their predictions to avoid underestimating transfer time with probability ratings greater than 50%. A probability rating greater than 50% would indicate a pessimistic prediction strategy as described in the paragraph above. In contrast to this hypotheses, many of the participants stated their predictions were optimistic in the consequence-information condition. However, this self-description did not match the participants' observed prediction behaviors reported in Section 7.3.2.

It was also hypothesized that when presented with likelihood-information, participants would center their internal models on the suggested adjustment, resulting in changes in the predicted time but a reported probability rating that is near 50%. Thus, participants were expected to self-report to pick an average case for their predictions. For the likelihood-information condition, most participants did self-report that their predictions were an average case (35 out of 79).

7.4.2 Self-reported Impact of Contextual Information on Prediction Behavior

Participants were presented with two statements, one for consequence-information (i.e., "The additional information provided about the patient condition or shift requirements of the crew impacted my estimate."), and one for likelihood-information (i.e., "The additional information provided about the traffic conditions or crew experience impacted my estimate."). Participants rated how much they agreed with each of these statements on a 5-point scale between Strongly Disagree and Strongly Agree. Table 7 shows the number of participants who responded in each category. The results suggest that both types of contextual information were considered to have an impact on prediction behavior, but likelihood-information was found to be more impactful than consequence-information. In addition, the results of this questionnaire provided further evidence that there was no difference between the two visualization types in terms of prediction behavior.

Table 7: Number of participants who reported that they agreed with a statement that the additional information provided by the two context conditions impacted their time predictions

		Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*	***Total***
Consequence	*Median-only*	2	6	6	35	30	79
	Boxplot	3	2	5	36	33	79
Likelihood	*Median-only*	0	0	0	27	52	79
	Boxplot	0	1	2	28	48	79

7.5 Discussion

The results of this study provide preliminary, but strong, evidence that contextual information is a factor that changes prediction behavior. Furthermore, the two framings of contextual information, consequence and likelihood, resulted in different types of prediction strategies. At a high level, these results suggest that when provided with decision support, in the form of visualizations of historical data, individuals are readily able to adjust their predictions based on information they have about the current situation, and the influence of this contextual information appears to be present both for visualizations that show variability information and those that do not.

In the no-context condition, participants were much more likely to choose the saliently presented central tendency point as hypothesized. Consequence-information changed the prediction behavior, with fewer participants choosing the central tendency, but with a relatively even distribution of predictions above and below the salient central tendency point. This was in contrast with the original hypothesis formed based on the results of the first experiment: consequence-information would result in predictions above the median indicating a potential bias to avoid underestimation. However, the likelihood-information appeared to strongly influence prediction behavior away from the central tendency point towards the time adjustment suggested by contextual information as was hypothesized.

It was surprising that visualization type (i.e., one with variability information and one without) did not have an effect on prediction behavior; its effect was limited to helping participants calibrate their confidence. This was in contrast with the first study, which found differences between the visualizations. One potential explanation for this contrast is that some of the datasets used in the first study were slightly skewed (see Appendix F for datasets) whereas the datasets used in the second study were created to be perfectly symmetric. Thus, in the first study, the effects may not have been solely due to the presence of variability information but also due to graphical features of the visualization. Evidence supporting the interaction between dataset skewness and visualization type was found in a third experiment (presented in Appendix I) conducted exactly as the second experiment, but with right-skewed historical datasets. This third study was run with 79 additional participants. Predictions made using the Median-only visualization, where skewness information was not visible, largely replicated those found in the second experiment. However, in the Boxplot visualization condition, participants were more likely to choose predictions in the direction of the skewness and tended to choose predictions that were further away from the central tendency point. Furthermore, the influence of the skewed distribution appeared to be strongest for the no-context condition, and least strong for the likelihood-information condition where participants' prediction behaviors appeared to be dominated by the context information provided.

The results in the Boxplot visualization condition for right-skewed distributions provides evidence that graphical features of a visualization (e.g., the length of the whiskers and the area of the box) may have significant influence on prediction behavior. In the Boxplot visualization, the median was visually represented as a thick straight line in the middle of the boxplot, and participants were provided an explanation that the median represented the 50th percentile point of the dataset. However, with skewed distributions the median would not represent the graphical center of the Boxplot visualization, leading to participants potentially adjusting their predictions away from this saliently denoted central tendency measure. Boxplots can be difficult to interpret because the area represented within each quartile does not map onto probability density, and alternative boxplot visualizations have been suggested that remedy this shortcoming (e.g., Benjamini, 1988). The framework proposed by this dissertation hypothesized that individuals base their predictions on their internal probability models; however, the results of the third experiment suggest that participants may have been relying on simpler graphical heuristics such as the length of the whiskers and boxes. Other visualization techniques that are better able to map graphical features to probability density, such as Violin plots (Hintze & Nelson, 1998) or quantile dot plots (Kay et al., 2016), may result in predictions that are more accurately tied to the underlying probability and should be examined in future work.

The second experiment also provided some insights into prediction strategies; however, strategy analysis is not straight-forward and the results are thus not conclusive. For consequence-information, there was greater variability in participants' prediction behaviors and strategies than had been anticipated. One possible explanation for the variability in responses is that participants were using the information provided in the contextual information stories as a basis for a simulation of possible outcomes which they used to adjust their predictions from the median. Hence, some participants may have been able to extrapolate potential changes in the underlying process that would occur because of the changes in consequence. For example, in a post-experiment feedback questionnaire, participants indicated that they felt that paramedic crews may attempt to rush the transfer for a critically-ill patient, or they may be slower in finishing a transfer when they were close to receiving overtime. The use of story building and mental simulation to assist in decision making and prediction is a common observation in Naturalistic Decision Making literature (Lipshitz, Klein, Orasanu, & Salas, 2001), and is likely less cognitively demanding then applying analytical methods for integrating the contextual information with the historical data. However, these simulations of possible outcomes are likely to be heavily influenced by the participants' own understanding of the patient transfer process and which factors they consider salient, resulting in greater inter-individual variability. Further research on how individual differences influence the interpretation of consequence-information is required.

In the likelihood-information condition, it was hypothesized that participants would adjust their internal models based on the likelihood-information in a way that shifted their internal probability models towards the time suggested by the contextual information. That is, participants would think that the adjustment would represent the central tendency of this specific case, in a way similar to how individual's knowledge about underlying base-rates influences their interpretation of uncertainty indicators (Wallsten et al., 1986; Weber & Hilton, 1990). Some participants confirmed this hypothesis by choosing predictions greater than the median (as suggested by the likelihood information) and stating that their predictions represented an “average” case (Median-only: 24 out of 79 participants; Boxplot: 19 out of 79). However, others’ behavior deviated from the hypothesis. Thus, it is possible that some participants adopted simpler heuristics for adjusting predictions instead of updating their internal probability models. The wording of the likelihood-information provided a simple recommendation (e.g., this dispatch team tends to complete the transfer 2 min slower than most crews) that the participants could act on. Since this information was provided as part of the experimental manipulation, participants may not have had any reason to doubt the validity of the information; the majority of the participants may have simply followed the suggestion without first interpreting the visualization and integrating the contextual information. Further research is required to examine whether the results will still hold when the likelihood-information is more uncertain (i.e., based on the individual’s own knowledge or information that they were provided through other sources).

7.6 Conclusion

In conclusion, this chapter showed that prediction behavior and strategies differ based on the type of contextual information and that users adjust their predictions when using historical data in response to contextual knowledge. Furthermore, the two different context types that were observed within Ornge influence predictions in different ways; but the reasons why these different contextual information change prediction behavior are still not well understood. Future studies should explore whether individuals are resorting to simple adjustment heuristics (i.e., anchor and adjust), and whether these heuristics are based on graphical features or based on mental simulation of what may occur in a given situation. Finally, the large variability in response strategies suggests that further interventions (e.g., training, interface and workflow changes) may be required to help users understand how to use historical data to predict transfer times, as it does not appear to be an intuitive process. For example, research has suggested that decision-aids may need to have explicit support and instructions for helping users integrate contextual information with the uncertainty information presented to them in a historical data visualization, as is supported by research on displaying meta-information in command and control interfaces (Pfautz et al., 2006). The next chapter provides an overall discussion of the experimental studies and relates them to field studies reported earlier.

Chapter 8

8 Overall Discussion of Experimental Studies

This chapter provides an overall discussion of the experimental studies described in Chapters 5 through 7 and relates them to the field studies and STP tool described in Chapters 2 through 4.

8.1 How do individuals use visualizations of historical data in making predictions?

Decision-aids displaying historical data have the potential to benefit the prediction of patient transport times to help improve dispatch decision making if users are able to interpret and use the historical data effectively. Since little prior research has examined how users interpret visualizations of historical data to make predictions, a series of experimental studies examined whether the framework proposed in Chapter 5 described the way the participants chose a predicted value. The framework hypothesized that the predicted value would be influenced by 1) the display format of the historical data and 2) information that is external to the cognitive-aid that is relevant to the current context of the prediction. The following sections will describe the type of prediction behavior found in the empirical studies and discuss the possible underlying processes that may have led to the exhibited prediction behavior.

8.1.1 What kind of prediction behavior do people exhibit when using visualizations of historical data?

One of the major contributions from the two experimental studies was the finding that the prediction of future values using visualizations of historical data is not merely a matter of visualization interpretation. Previous studies, such as Ibrekk and Morgan (1987) and Edwards et al. (2012), largely focused on the ability of individuals to produce estimates of a specific measure of interest, such as the mean of the data. These studies demonstrated that displays that directly present the measure of interest result in the best performance. Furthermore, the little research available on future variable predictions using uncertainty visualizations (e.g., Nadav-Greenberg & Joslyn, 2009) found that predicted values are not significantly different than the point-estimates provided in the display. Thus, it may be expected that when provided with a visualization displaying historical data, participants may simply choose the central tendency of the presented data as their prediction. In both of the empirical studies reported in this dissertation, a measure of central tendency was saliently displayed and could be easily selected by participants using the interactive interfaces provided, allowing participants to easily adopt this strategy.

However, the first experiment showed participants may have adjusted their predictions based on how the predictions would be used. That is, instead of choosing the saliently presented central tendency point, most participants overestimated the time, potentially to provide a safety margin. The role of context was explicitly tested in the second experiment. Participants adjusted their predictions toward the value suggested by the likelihood-information, while how participants adjusted their predictions when given consequence-information varied with no specific result. Taken together, the results indicate that the task of predicting future values using historical data visualizations, which is understudied, is conceptually different from a graph interpretation task (e.g., Ibrekk & Morgan, 1987). It appears that context can play a large role. One limitation of the empirical studies presented in this dissertation is that the structure of the task used always favored overestimations (i.e., the perceived costs of underestimating time for rover scheduling appeared to be greater than that of overestimating; the likelihood- and consequence-information both provided adjustment values that were slower than the median time). Future research should also manipulate contextual information favoring underestimations.

The influence of contextual information on prediction behavior was found both when the visualization provided variability information and when it did not. While the visualizations explored in the two experiments represented varying amounts of uncertainty aggregation (i.e., changes in the amount of variability information), the results provided evidence that graphical features of the visualizations may have been the primary source of the different prediction behaviors observed rather than the amount of variability information provided. Further work is required to explore how different features affect prediction behavior; broader theories of how individuals interpret graphics such as the one by Pinker (1990) may be useful for this purpose. In general, previous reviews of risk communication techniques using visualizations (Lipkus & Hollands, 1999; Spiegelhalter, Pearson, & Short, 2011) have highlighted a lack of research connecting visualization design to the structure of the task being investigated. This dissertation helped narrow this gap through the investigation of one highly understudied task, the prediction of a future value of a variable.

It is interesting to note the contrast between the tendency of the participants in the first experiment to overestimate predicted times and the tendency of Ornge's planners to underestimate the time to definitive care (compared to actual transport times) as was reported in Chapter 3. The underestimates by the planners at times were due to a lack of consideration for parts of the transfer process, whereas the task used in the experiment was much simpler. Other pre-hospital emergency care providers have also been found to make underestimations of transfer times (Propp & Rosenberg, 1991). In the experiments, participants were generating predictions supported by decision-aid presenting historical data, while Ornge's dispatchers were generating their predictions largely based on their own experiences and simple heuristics (Giang, Donmez, Fatahi, et al., 2014). The more subjective, unsupported (i.e., without the aid of historical data) time prediction processes have been found to be prone to underestimations (Halkjelsvik & Jørgensen, 2012). Thus, the simplicity of the tasks used in the experiments is a limitation and can only uncover a small aspect of how actual planners' predictions may be influenced by a decision-aid. It is likely that with the historical data visualizations used in the experimental studies, the overestimations were a conscious choice by the participant, while the Ornge dispatchers may not have been aware of their bias towards underestimation, as previous studies have found that underestimations in the prediction of future time durations can be attributed to memory errors of how long events have taken in the past (Roy & Christenfeld, 2007; Roy, Mitten, & Christenfeld, 2008).

Furthermore, the participants in both experiments were able to correctly adjust their confidence in their predictions when the underlying datasets had larger variability and when historical variability information was presented. This result reaffirms previous studies on the benefits of uncertainty information on improving decision making (Joslyn & LeClerc, 2013; Nadav-Greenberg & Joslyn, 2009; Savelli & Joslyn, 2013). These calibrations in confidence can lead to better informed decisions throughout the dispatch process. Thus, providing additional variability information not only can remind the dispatchers of travel intervals that are more susceptible to disruptions, it can also help them consider the reliability of their final predicted values. However, further validation in the application domain is still required. Future work with actual medical dispatchers is required to validate whether more accurate calibration of prediction confidence will lead to improvements in dispatch decision making. Furthermore, transport planners communicating their confidence in their predicted time to definitive care estimates to other dispatch decision makers should be examined as a possible reason why historical variability information may lead to better dispatch decision making.

8.1.2 How do people generate predictions when using visualizations of historical data?

Chapter 5 proposed that individuals make predictions that are based on their understanding of the underlying distribution of the process represented in the historical data. That is, individuals interpret historical data visualizations to generate an internal probability model first (e.g., transport times); they then select a point within this internal distribution as their predicted value. While the results of the two experiments presented in this dissertation provide evidence that both display format and contextual information have statistically significant effects on prediction behavior, the results did not provide any clear support for the hypothesized prediction process, in particular, the generation of an internal model.

The first experiment showed that participants deviated further from the central tendency when the magnitude of the variability presented to them was larger. This result provides evidence that participants chose a prediction that scaled with the underlying distribution of the process. However, it is unclear whether participants were actually generating an internal probability model, or whether they were simply using graphical cues from the visualizations to help them locate a desired point, as discussed earlier in Chapter 7. It is likely that interpreting the entire probability distribution before choosing a prediction is a time consuming and cognitively demanding task and would only be attempted when information about the entire distribution is needed. In the Tak et al. (2014, 2015) studies, participants' internal models were queried explicitly which allowed the authors to test whether their participants were able to accurately interpret the probabilities represented by the visualization. However, if participants were only asked to produce a single piece of information, such as the predicted time in the studies presented in this dissertation, the use of simpler graphical heuristics may require less effort. The underlying processes that drove the changes observed with the different factors explored requires further investigation.

8.2 Recommendations for Ornge's Short-term Planning Tool

Based on the results of the experimental studies the following recommendations were also generated for the implementation of the STP tool for Ornge dispatchers. First, when dispatchers want to adjust their predictions away from the suggested historical central tendency, they should be asked to provide justification. This justification would serve as record keeping and also encourage the dispatcher to think about what contextual information justifies the adjustment. Second, dispatchers should be reminded to generate accurate time estimates rather than potentially building safety room into their predictions. The results of the dissertation provide evidence that non-dispatchers may consider contextual factors about how predictions are used in order to adjust their predictions when using decision-aids with historical data. However, Ornge's dispatchers communicate their predictions to other dispatchers to support further dispatch decision making and having additional flex room may bias subsequent decision-makers. In addition to reminders and training to produce accurate predictions, the STP tool should use the algorithm outputs as default predictions, which may mitigate the bias towards including consequence context information in the predictions. Finally, dispatchers can communicate their confidence in their predictions in addition to their predictions.

Chapter 9

9 Conclusion and Future Work

In conclusion, this dissertation examined the following research questions:

- Practical Research Questions: What is the role of patient transport time predictions in medical dispatch, and can information about the variability of historical transport times improve these predictions?

- Theoretical Research Questions: How are commonly used historical data visualizations (e.g., boxplots) interpreted by individuals to support the prediction of future values of a variable, and what factors influence these predictions?

In addition to statistical analysis of historical Ornge data, two field studies and two experimental studies were conducted. The findings suggest that:

- Dispatchers make use of patient transport time predictions to support scheduling, logistics, and dispatch decisions, however dispatchers tend to underestimate transport times.

- Dispatchers may benefit from having historical variability information to help gauge the reliability of a decision-aid. This conclusion is supported by the experimental evidence: individuals adjust their confidence in their time predictions based on the magnitude of the variability in the dataset shown.

- Contextual factors (e.g., weather, patient severity) are considered by dispatchers when making dispatch decisions. Experiments also showed that contextual factors influence prediction behavior and that different types of contextual information observed in the field (i.e., those that change the likelihood of different outcomes, and those that change the consequence of different outcomes) affect prediction behavior in different ways.

- The prediction of future values, using historical data, is influenced by the visualization used to present the historical data, but prediction behavior is not solely dependent on how well individuals are able to interpret the visualization; contextual factors as discussed above influence the choice of a predicted value.

- The various display formats used to present historical data change the type of information individuals have access to, which result in different types of prediction strategies.

- Human dispatchers assisted with a decision-aid built using historical transport time data is a promising method to produce more accurate predictions. Although this dissertation does not explicitly test whether the proposed tool enhanced prediction accuracy, it provides support to this claim by the literature reviewed as well as the findings of the field and experimental studies.

These findings were used in the design of a Short-term Planning tool to help Ornge's dispatchers with generating transport time predictions. This tool is currently being adopted by Ornge.

Limitations specific to the field and empirical studies have been listed in earlier chapters where these studies are discussed in detail. The below list reiterates some of the major ones of these limitations:

- The field studies were conducted at a single large-scale air ambulance service in Canada, and further research is needed to examine whether the findings are generalizable to other medical transportation systems.

- The use of a general population, recruited online, instead of trained dispatchers in the empirical studies limit the generalization of the empirical findings to the medical dispatch domain. Further research should examine the prediction behavior and strategies of Ornge's dispatchers with the visualizations of historical data.

- A highly controlled, artificial time prediction task was used in the empirical studies, rather than a dynamic, context-specific dispatch decision making task. Further work to extend the findings of this dissertation to more complex decision tasks using microworld simulations or other simulation studies can help address this limitation.

Suggestions for future work have been proposed in earlier chapters. The below list reiterates some of the major lines of research that can be pursued in the future building on the contributions of this dissertation:

- The influence of the decision-aid proposed in this dissertation on team cognition can be investigated. The introduction of uncertainty information, using visualizations of historical data, may help improve communication and team decision-making.

- Although this dissertation proposed that individual differences may play a role in how individuals use historical data visualizations, further research is needed to explore this hypothesis.

- As noted earlier, the findings of the field and empirical studies were used in the design of a Short-term Planning tool to help Ornge's dispatchers with generating transport time predictions. This tool is currently being adopted by Ornge. Future research should evaluate the effectiveness of this tool in operation.

- Judgement analysis (Cooksey, 1996), and its extensions (e.g., Human-Auotmated Judge Learning (HAJL), Bass & Pritchett, 2008) can be adopted to provide further insight into the prediction behavior of dispatchers when using historical data decision-aids. In particular, HAJL may be useful to understand what visual cues within a visualization influence prediction behavior.

References

Bass, E. J., Baumgart, L. A., & Shepley, K. K. (2013). The effect of information analysis automation display content on human judgment performance in noisy environments. *Journal of Cognitive Engineering and Decision Making*, *7*(1), 49–65.

Bass, E. J., & Pritchett, A. R. (2008). Human-automated judge learning: A methodology for examining human interaction with information analysis automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, *38*(4), 759–776.

Baumgart, L. A., Bass, E. J., Voss, J. D., & Lyman, J. A. (2015). Effect of pooled comparative information on judgments of quality. *IEEE Transactions on Human-Machine Systems*, *45*(6), 773–781.

Benjamini, Y. (1988). Opening the box of a boxplot. *The American Statistician*, *42*(4), 257–262.

Beyer, H., & Holtzblatt, K. (1998). *Contextual Design: Defining Customer-Centered Systems*. San Fransico, CA: Morgan Kaufmann Publishers.

Bisantz, A. M., Cao, D., Jenkins, M., Pennathur, P. R., Farry, M., Roth, E., … Pfautz, J. (2011). Comparing uncertainty visualizations for a dynamic decision-making task. *Journal of Cognitive Engineering and Decision Making*, *5*(3), 277–293.

Bisantz, A. M., D'Arcy, J.-F., Kerker, D., Hegde, S., Guan, P., Voshell, M., & Kilgore, R. (2014). Maps, space-time cubes, and meta-information for understanding path information: a comparative analysis. In *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting* (pp. 370–374). Chicago, IL.

Bisantz, A. M., Marsiglio, S. S., & Munch, J. (2005). Displaying uncertainty: investigating the effects of display format and specificity. *Human Factors*, *47*(4), 777–796.

Bisantz, A. M., Stone, R. T., Pfautz, J., Roth, E., Nagy, A. L., & Thomas, G. (2009). Visual representations of meta-information. *Journal of Cognitive Engineering and Decision Making*, *3*(1), 67–91.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101.

Buehler, R., Griffin, D., & MacDonald, H. (1997). The role of motivated reasoning in optimistic time predictions. *Personality and Social Psychology Bulletin*, *23*(3), 238–247.

Buehler, R., Griffin, D., & Peetz, J. (2010). The planning fallacy: cognitive, motivational, and social origins. In M. Zanna & J. Olson (Eds.), *Advances in Experimental Social Psychology* (1st ed., Vol. 43, pp. 1–62). Burlington: Academic Press.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": why people underestimate their task completion times. *Journal of Personality and Social Psychology*, *67*(3), 366–381.

Burt, C. D. B., & Kemp, S. (1994). Construction of activity duration and time management potential. *Applied Cognitive Psychology*, *8*(2), 155–168.

Burton, O., & Mccarley, J. S. (2017). Visualization of uncertainty aids spatial judgments but fails to improve metacognitive efficiency. In *Proceedings of the Human Factors and Ergonomics Society 61st Annual Meeting* (pp. 1390–1393). Austin, TX.

Byram, S. J. (1997). Cognitive and motivational factors influencing time prediction. *Journal of Experimental Psychology: Applied*, *3*(3), 216–239.

Chow, R., & Vicente, K. J. (2002). A field study of emergency ambulance dispatching: Implications for decision support. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 313–317). Baltimore, MD.

Clawson, J. J., & Dernocoeur, K. B. (1998). *Principles of Emergency Medical Dispatch* (2nd ed.). Salt Lake City: The National Academy of Emergency Medical Dispatch, Priority Press.

Cooksey, R. W. (1996). *Judgement analysis: Theory, methods, and applications*. New York: Academic Press.

Correll, M., & Gleicher, M. (2014). Error bars considered harmful: exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 2142–2151.

Cuellar, A. E., & Gertler, P. J. (2003). Trends in hospital consolidation: the formation of local systems. *Health Affairs*, *22*(6), 77–87.

Dawes, R., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *59*(2), 257–272.

Dodge, K. A., & Mandel, A. D. (2012). Building evidence for evidence-based policy making. *Criminology & Public Policy*, *11*(3), 525–534.

Edwards, J. A., Snyder, F. J., Allen, P. M., Makinson, K. A., & Hamby, D. M. (2012). Decision making for risk management: a comparison of graphical methods for presenting quantitative uncertainty. *Risk Analysis*, *32*(12), 2055–2070.

Fatahi, A. (2013). *Air versus Land Vehicle Decisions for Interfacility Air Medical Transport*. University of Toronto.

Fatahi, A., Donmez, B., Ahghari, M., & MacDonald, R. D. (2012). Air versus ground vehicle decisions for interfacility air medical transport. In *Proceedings of the 2012 Industrial and Systems Engineering Research Conference*. Orlando, FL.

Ferno. (2018). ACETech. Retrieved June 15, 2018, from http://www.fernoacetech.com/

Finger, R., & Bisantz, A. M. (2002). Utilizing graphical formats to convey uncertainty in a decision-making task. *Theoretical Issues in Ergonomics Science*, *3*(1), 1–25.

Fletcher, K. A., & Bedwell, W. L. (2017). Cognitive aids in emergency medical services. In J. R. Keebler, E. H. Lazzara, & P. Misasi (Eds.), *Human Factors and Ergonomics of Prehospital Emergency Care* (pp. 123–139). Boca Raton, Fl: CRC Press.

Flyvbjerg, B., Holm, S., & Buhl, S. (2002). Cost underestimation in public works projects: error or lie? *Journal of the American Planning Association*, *68*(3), 279–295.

Furniss, D., & Blandford, A. (2006). Understanding emergency medical dispatch in terms of distributed cognition: A case study. *Ergonomics*, *49*(12–13), 1174–1203.

Giang, W. C. W., & Donmez, B. (2015). Interpreting Visualizations of Historical Variability for Estimating Future Events Interpreting Visualizations of Historical Variability for Estimating Future Events. *Proceedings of the International Conference on Naturalistic Decision Making 2015*, (1), 4.

Giang, W. C. W., Donmez, B., Ahghari, M., & MacDonald, R. D. (2014). The impact of precipitation on land interfacility transport times. *Prehospital and Disaster Medicine*, *29*(6), 593–599.

Giang, W. C. W., Donmez, B., Fatahi, A., Ahghari, M., & MacDonald, R. D. (2014). Supporting air versus ground vehicle decisions for interfacility medical transport using historical data. *IEEE Transactions on Human-Machine Systems*, *44*(1), 55–65.

Giang, W. C. W., Donmez, B., Zakir, A., Ahghari, M., & MacDonald, R. D. (2015). Supporting dispatch decisions in interfacility medical transfers: Understanding the roles of uncertainty and reliability. In *Proceedings of the 4th International Symposium on Human Factors and Ergonomics in Health Care* (pp. 174–178). Baltimore, MD.

Giang, W. C. W., Hui, L., Donmez, B., Ahghari, M., & MacDonald, R. D. (2016). Dispatch decision making in an air medical transport system. In *Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting* (pp. 259–263). Washington, DC.

Greis, M., Ohler, T., Henze, N., & Schmidt, A. (2015). Investigating representation alternatives for communicating uncertainty to non-experts. In *INTERACT 2015* (pp. 256–263). Bamberg, Germany.

Grounds, M. A., & Joslyn, S. L. (2018). Communicating weather forecast uncertainty: Do individual differences matter? *Journal of Experimental Psychology: Applied*, *24*(1), 18–33.

Grounds, M. A., Joslyn, S. L., & Otsuka, K. (2017). Probabilistic interval forecasts: an individual differences approach to understanding forecast communication. *Advances in Meteorology*, *2017*, 1–19.

Grove, W. M., & Lloyd, M. (2006). Meehl's contribution to clinical versus statistical prediction. *Journal of Abnormal Psychology*, *115*(2), 192–194.

Gschwandtnei, T., Bögl, M., Federico, P., & Miksch, S. (2016). Visual encodings of temporal uncertainty: A comparative user study. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 539–548.

Guarino, S., Pfautz, J., Cox, Z., & Roth, E. (2009). Modeling human reasoning about meta-information. *International Journal of Approximate Reasoning*, *50*(3), 437–449.

Guessous, Y., Aron, M., Bhouri, N., & Cohen, S. (2014). Estimating travel time distribution under different traffic conditions. *Transportation Research Procedia*, *3*, 339–348.

Halkjelsvik, T., & Jørgensen, M. (2012). From origami to software development: a review of studies on judgment-based predictions of performance time. *Psychological Bulletin*, *138*(2), 238–271.

Harenčárová, H. (2015). Structured analysis of Critical Decision Method data – emergency medicine case study. *Human Affairs*, *25*(4), 443–459.

Harrington, D. T., Connolly, M., Biffl, W. L., Majercik, S. D., & Cioffi, W. G. (2005). Transfer times to definitive care facilities are too long. *Annals of Surgery*, *241*(6), 961–968.

Herdener, N., Wickens, C. D., Clegg, B. A., & Smith, C. A. P. (2015). Mapping spatial uncertainty in prediction. In *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting* (pp. 140–144). Los Angeles, CA.

Hintze, J. L., & Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *The American Statistician*, *52*(2), 181–184.

Hullman, J., Resnick, P., & Adar, E. (2015). Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PLoS ONE*, *10*(11), 1–25.

Ibrekk, H., & Morgan, M. G. (1987). Graphical communication of uncertain quantities to nontechnical people. *Risk Analysis*, *7*(4), 519–529.

Institute of Medicine. (2007). *Emergency Medical Services: At the Crossroads*. Washington, DC: The National Academies Press.

Joslyn, S. L., & Grounds, M. A. (2015). The use of uncertainty forecasts in complex decision tasks and various weather conditions. *Journal of Experimental Psychology: Applied*, *21*(4), 407–417.

Joslyn, S. L., & LeClerc, J. (2013). Decisions with uncertainty: the glass half full. *Current Directions in Psychological Science*, *22*(4), 308–315.

Joslyn, S. L., & LeClerc, J. E. (2012). Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *Journal of Experimental Psychology. Applied*, *18*(1), 126–140.

Joslyn, S. L., Nadav-Greenberg, L., & Nichols, R. M. (2009). Probability of precipitation: assessment and enhancement of end-user understanding. *Bulletin of the American Meteorological Society*, *90*(2), 185–193.

Kahneman, D., & Tversky, A. (1982a). Intuitive prediction: biases and corrective procedures. In

D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement Under Uncertainty: Heuristics and Biases* (pp. 414–421). London: Cambridge University Press.

Kahneman, D., & Tversky, A. (1982b). Variants of uncertainty. *Cognition*, *11*(2), 143–157.

Kanten, A. B. (2011). The effect of construal level on predictions of task duration. *Journal of Experimental Social Psychology*, *47*(6), 1037–1047.

Kay, M., Kola, T., Hullman, J. R., & Munson, S. A. (2016). When (ish) is my bus? User-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5092–5103). San Jose, CA.

Khamooshi, H., & Cioffi, D. F. (2013). Uncertainty in task duration and cost estimates: Fusion of probabilistic forecasts and deterministic scheduling. *Journal of Construction Engineering and Management*, *139*(5), 488–497.

Kirlik, A. (2007). Lessons learned from the design of the decision support system used in the hurricane katrina evacuation decision. In *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting* (pp. 253–257). Baltimore, MD.

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man and Cybernetics*, *19*(3), 462–472.

Koch, C. J., & Kleinmann, M. (2002). A stitch in time saves nine: behavioural decision-making explanations for time management problems. *European Journal of Work and Organizational Psychology*, *11*(2), 199–217.

König, C. J. (2005). Anchors distort estimates of expected duration. *Psychological Reports*, *96*(2), 253–256.

Latitude Technologies. (2018). Latitude Technologies. Retrieved June 15, 2018, from http://latitudetech.com/

Lerner, E., Billittier, A., Sikora, J., & Moscati, R. (1999). Use of a geographic information system to determine appropriate means of trauma patient transport. *Academic Emergency Medicine*, *6*(11), 1127–1133.

Lipkus, I. M., & Hollands, J. G. (1999). The visual communication of risk. *Journal of the National Cancer Institute. Monographs*, (25), 149–163.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Focus article: taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, *14*(5), 331–352.

Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: a naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes*, *69*(2), 149–163.

Mannes, A. E., & Moore, D. A. (2013). A behavioral demonstration of overconfidence in judgment. *Psychological Science*, *24*(7), 1190–1197.

Martin, T. (2006). *Aeromedical Transportation: A Clinical Guide* (2nd ed.). Aldershot: Ashgate Publishing, Ltd., 2006.

Maule, A. (1997). Strategies for adapting to time pressure. In R. Flin, M. Salas, M. Strub, & I. Martin (Eds.), *Decision Making Under Stress: Emerging Themes and Applications* (pp. 271–279). Brookfield, VT: Ashgate Publishing Company.

McQueary, B., Krause, L., Santos, E., Wang, H., & Zhao, Q. (2004). Modeling, analysis and visualization of uncertainty in the battlespace. In *Proceedings of the 16th IEEE International Conference on Tools with Artifical Intelligence* (pp. 782–783). Boca Raton, Fl.

Meehl, P. (1954). *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis, MN: University of Minnesota Press.

Miller, S., Kirlik, A., Kosorukoff, A., & Tsai, J. (2008). Supporting joint human-computer judgment under uncertainty. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 408–412).

Nadav-Greenberg, L., & Joslyn, S. L. (2009). Uncertainty forecasts improve decision making among nonexperts. *Journal of Cognitive Engineering and Decision Making*, *3*(3), 209–227.

National Research Council. (1996). *Understanding Risk: Informing Decisions in a Democratic Society*. (P. C. Stern & H. V Fineberg, Eds.). Washington, DC: The National Academies Press.

Neyedli, H. F., Hollands, J. G., & Jamieson, G. A. (2011). Beyond identity: incorporating system reliability information into an automated combat identification system. *Human Factors*, *53*(4), 338–355.

Office of the Chief Coroner for Ontario. (2013). *Review of Ornge Air Ambulance Transport Related Deaths*. Retrieved from https://www.mcscs.jus.gov.on.ca/sites/default/files/content/mcscs/docs/ec163557.pdf

Ornge. (2017a). About us. Retrieved November 5, 2017, from https://www.ornge.ca/about

Ornge. (2017b). Types of response. Retrieved November 4, 2017, from https://www.ornge.ca/healthcare/occ-(1)/response-type

Palmer, E. M. (2017). Cognitive factors in emergency medical services. In J. R. Keebler, E. H. Lazzara, & P. Misasi (Eds.), *Human Factors and Ergonomics of Prehospital Emergency Care* (pp. 13–25). Boca Raton, Fl.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. New York, NY: Cambridge University Press.

Pfautz, J., Roth, E., Bisantz, A. M., Thomas-Myers, G., Llinas, J., & Fouse, A. (2006). *The role of meta-information in C2 decision-support systems*. Cambridge, MA.

Pfeffer, J., & Sutton, R. I. (2006). Evidence-based management. *Harvard Business Review*, *84*(1), 62–74, 133.

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial Intelligence and the Future of Testing* (1st ed., pp. 73–126). New York, NY: Psychology Press.

Propp, D. A., & Rosenberg, C. A. (1991). A comparison of prehospital estimated time of arrival and actual time of arrival to an emergency department. *The American Journal of Emergency Medicine*, *9*(4), 301–303.

Pugh, A. J., Wickens, C. D., Herdener, N., Clegg, B. A., & Smith, C. A. P. (2017). Effect of visualization on spatial trajectory prediction under uncertainty. In *Proceedings of the Human Factors and Ergonomics Society 61st Annual Meeting* (pp. 297–301). Austin, TX.

Reiss, J. (2004). Evidence-based economics. *Analyse and Kritik*, *26*, 346–363.

Roy, M. M., & Christenfeld, N. J. S. (2007). Bias in memory predicts bias in estimation of future task duration. *Memory & Cognition*, *35*(3), 557–564.

Roy, M. M., Christenfeld, N. J. S., & McKenzie, C. R. M. (2005). Underestimating the duration of future events: memory incorrectly used or memory bias? *Psychological Bulletin*, *131*(5), 738–756.

Roy, M. M., Mitten, S. T., & Christenfeld, N. J. S. (2008). Correcting memory improves accuracy of predicted task duration. *Journal of Experimental Psychology. Applied*, *14*(3), 266–275.

Ruginski, I. T., Boone, A. P., Padilla, L. M., Liu, L., Heydari, N., Kramer, H. S., … Creem-Regehr, S. H. (2016). Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition & Computation*, *16*(2), 154–172.

Sackett, D., Rosenberg, W., Gray, J., Haynes, R. B., & Richardson, W. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, *312*(7023), 71–72.

Savelli, S., & Joslyn, S. L. (2013). The advantages of predictive interval forecasts for non-expert users and the impact of visualizations, *27*(4), 527–541.

Singh, J. M., & MacDonald, R. D. (2009). Pro/con debate: do the benefits of regionalized critical care delivery outweigh the risks of interfacility patient transport? *Critical Care*, *13*(4), 219.

Slack, D., Koenig, K., & Bouley, D. (1995). Paramedic accuracy in estimated time of arrival: significance in the managed care environment. *Academic Emergency Medicine*, *2*(10), 943–944.

Smith, J. S., Smith, B. J., Pletcher, S. E., Swope, G. E., & Kunst, D. (1993). When is air medical service faster than ground transportation? *Air Medical Journal*, *12*(8), 258–261.

Sockolow, P. S., Ms, D., Yang, Y., Bass, E. J., Rn, K. H. B., Holmberg, A., … York, N. (2017). Data visualization of home care admission nurses' decision-making. In *AMIA Annual*

Symposium Proceedings (pp. 1597–1606).

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science*, *333*(6048), 1393–1400.

Strauss, A., & Corbin, J. (1994). *Grounded theory methodology*. (N. K. Denzin & Y. S. Lincoln, Eds.), *Handbook of qualitative research* (17th ed.). Thousand Oaks CA: Sage Publications.

Svenson, J. E., O'Connor, J. E., & Lindsay, M. B. (2006). Is air transport faster? A comparison of air versus ground transport times for interfacility transfers in a regional referral system. *Air Medical Journal*, *25*(4), 170–172.

Tak, S., Toet, A., & van Erp, J. (2014). The perception of visual uncertainty representation by non-experts. *IEEE Transactions on Visualization and Computer Graphics*, *20*(6), 935–943.

Tak, S., Toet, A., & van Erp, J. (2015). Public understanding of visual representations of uncertainty in temperature forecasts. *Journal of Cognitive Engineering and Decision Making*, *9*(3), 241–262.

Thomas, K. E., & Handley, S. J. (2008). Anchoring in time estimation. *Acta Psychologica*, *127*(1), 24–29.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*(4157), 1124–1131.

Visschers, V. H. M., Meertens, R. M., Passchier, W. W. F., & de Vries, N. N. K. (2009). Probability information in risk communication: a review of the research literature. *Risk Analysis*, *29*(2), 267–287.

Walker, W. E., Harremoes, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B. A., Janssen, P., & Krayer von Krauss, M. P. (2003). A conceptual basis for uncertainty management. *Integrated Assessment*, *4*(1), 5–17.

Wallsten, T. S., Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language*, *25*(5), 571–587.

Weber, E. U., & Hilton, D. J. (1990). Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 781–789.

Windschitl, P. D., & Weber, E. U. (1999). The interpretation of "likely" depends on the context, but "70%" is 70%--right? The influence of associative processes on perceived certainty. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(6), 1514–1533.

Wong, B. L. W. (2003). Critical Decision Method data analysis. In D. Diaper & N. A. Stanton (Eds.), *The Handbook of Task Analysis for Human-Computer Interaction* (pp. 327–346). Mahwah: Lawrence Erlbaum Associates.

Wong, B. L. W., & Blandford, A. (2002). Analysing ambulance dispatcher decision making:

Trialing emergent themes analysis. In *Proceedings of the HF2002 Human Factors Conference: Design for the whole person: integrating physical, cognitive and social aspects*. Melbourne, Australia.

Wong, B. L. W., Hayes, J., & Moore, T. (2005). What makes emergency ambulance command and control complex? In C. Johnson (Ed.), *2nd Workshop on Complexity in Design and Engineering* (pp. 106–116). Glasgow, Scotland.

Wong, B. L. W., O'Hare, D., & Sallis, P. J. (1996). A goal-oriented approach for designing decision support displays in dynamic environments. In *Proceedings of the Sixth Australian Conference on Computer Human Interaction* (pp. 78–85). Hamilton, Australia.

Wong, B. L. W., Sallis, P. J., O'Hare, D., & Wong, William B L, Sallis, P.J., O'Hare, D. (1997). Eliciting information portrayal requirements: experiences with the Critical Decision Method. In H. Thimbleby, B. O'Conaill, & P. Thomas (Eds.), *Proceedings of HCI' 97 Conference of the British Computer Society Special Interest Group on Human-Computer Interaction* (pp. 397–415). Bristol, UK: Springer London.

Yaniv, I., & Hogarth, R. (1993). Judgmental versus statistical prediction: information asymmetry and combination rules. *Psychological Science*, *4*(1), 58–62.

Appendix A – Consent Documents for Field Studies

1. <u>Contextual Inquiry Study</u>

UNIVERSITY OF TORONTO

<u>INFORMED CONSENT BY SUBJECTS TO PARTICIPATE IN A RESEARCH EXPERIMENT</u>

Project: Short-term planning decision support for emergency medical transport

I have read the information presented in the information letter about a study being conducted by Wayne Giang and Lavinia Hui under the supervision of Professor Birsen Donmez of the Department of Mechanical and Industrial Engineering at the University of Toronto. I have had the opportunity to ask any questions related to this study, to receive satisfactory answers to my questions, and any additional details I wanted.

I am aware that I may withdraw my study participation at any time without penalty by advising the researcher. I am also aware that I can decline to answer any questions during the interview without any consequences.

This project has been reviewed by, and received ethics clearance through, the Office of Research Ethics at the University of Toronto. I was informed that if I have any comments or concerns resulting from my participation in this study, I may contact the Office of Research Ethics at ethics.review@utoronto.ca or 416-946-3273.

	Please Circle One	
With full knowledge of all foregoing, I agree, of my own free will, to participate in this study.	**YES**	**NO**
I agree to let my conversation during the study be directly quoted, anonymously, in presentation of the research results.	**YES**	**NO**

Participant Name: ____ (Please print)

Participant Signature: ____

Witness Name: ____ (Please print)

Witness Signature: ____

Date: ____

2. CDM Study 1 (Transport Planners)

UNIVERSITY OF TORONTO

INFORMED CONSENT BY SUBJECTS TO PARTICIPATE IN A RESEARCH EXPERIMENT

Project: Short-term planning decision support for emergency medical transport

I have read the information presented in the information letter about a study being conducted by Wayne Giang and Lavinia Hui under the supervision of Professor Birsen Donmez of the Department of Mechanical and Industrial Engineering at the University of Toronto. I have had the opportunity to ask any questions related to this study, to receive satisfactory answers to my questions, and any additional details I wanted.

I am aware that I may withdraw my study participation at any time without penalty by advising the researcher. I am also aware that I can decline to answer any questions during the interview without any consequences.

This project has been reviewed by, and received ethics clearance through, the Office of Research Ethics at the University of Toronto. I was informed that if I have any comments or concerns resulting from my participation in this study, I may contact the Office of Research Ethics at ethics.review@utoronto.ca or 416-946-3273.

	Please Circle One	
With full knowledge of all foregoing, I agree, of my own free will, to participate in this study.	**YES**	**NO**
I agree to let my conversation during the study be directly quoted, anonymously, in presentation of the research results.	**YES**	**NO**

Participant Name: ______________________________ (Please print)

Participant Signature: ______________________________

Witness Name: ______________________________ (Please print)

Witness Signature: ______________________________

Date: ______________________________

3. CDM Study 2 (OMs and TMPs)

UNIVERSITY OF TORONTO

INFORMED CONSENT BY SUBJECTS TO PARTICIPATE IN A RESEARCH EXPERIMENT

Project: Short-term planning decision support for emergency medical transport

I have read the information presented in the information letter about a study being conducted by Wayne Giang under the supervision of Professor Birsen Donmez of the Department of Mechanical and Industrial Engineering at the University of Toronto. I have had the opportunity to ask any questions related to this study, to receive satisfactory answers to my questions, and any additional details I wanted.

I am aware that the data collected from this study (e.g., the audio recordings, interview transcripts, and notes) will be stored indefinitely in a secure server within Prof. Donmez's lab at the University of Toronto.

I am aware that I may withdraw my study participation at any time without penalty by advising the researcher. I am also aware that I can decline to answer any questions during the interview without any consequences. I was also informed that the researchers will contact me after my withdrawal from the study. I may choose to provide permission for the researchers to use the data collected up to the point of withdrawal as part of their analysis. If I choose to decline, the data will be deleted immediately.

I was informed that if I have any comments or concerns resulting from my participation in this study, I may contact the Office of Research Ethics at ethics.review@utoronto.ca or 416-946-3273. I am also aware that the Human Research Ethics Program may have confidential access to data collected in this study to help ensure participant protection procedures are followed.

	Please Circle One	
With full knowledge of all foregoing, I agree, of my own free will, to participate in this study.	**YES**	**NO**
I agree to let my conversation during the study be directly quoted, anonymously, in presentation of the research results.	**YES**	**NO**
I agree to be recorded electronically using a digital voice recorder.	**YES**	**NO**

Participant Name: ____ (Please print)

Participant Signature: ____

Witness Name: ____ (Please print)

Witness Signature: ____

Date: ____

Appendix B – Dispatch Case Studies from Field Study 2 in Chapter 3

Case 1 (Role – TMP): Level of care/Active Management case.

A woman in a nursing station in northern Ontario who is reportedly to be in labour at 34 weeks. TMP is informed that it is her first natural birth and that she was leaking amniotic fluid. TMP sets the level of care at Advanced, which was the fastest crew available. Ornge paramedics arrive at nursing station, report to OCC that the patient was starting to have back pains and her cervix was dilating. TMP decides that the paramedics should stay at the nursing station for now and consults with pediatric TMP. Pediatric TMP suggests to the nursing station that they request a neonatal team from Winnipeg, which the nursing station does. TMP changes the level of care to critical, as a critical care crew was now available and is also dispatched to the nursing station. Critical care paramedic (CCP) crew arrives, followed closely by the neonatal team who concludes that the patient was not actually in active labour. Neonatal team transports patient to Winnipeg along with Ornge's CCP medic.

Major Decisions: 1) Initial level of care decision. 2) Decision to transport patient or stay at nursing station once ACP crew arrived. 3) Decision to send a CCP crew.

Sources of Uncertainty: 1) Patient's current condition. 2) Timeline for the patient's delivery. 3) Care strategy for premature baby. 3) Whether the level of care currently allocated was sufficient once the case became more dynamic. 4) Operational timelines (e.g., neonatal team, Ornge's crews).

Case 2 (Role – TMP): Patient Triage case.

CCP crew aboard a helicopter was close to arriving at Pembroke for an emergent interfacility transfer of a patient who had ruptured her membrane (patient 1). Scene call came in for a young woman who was in respiratory distress near Algonquin park (patient 2). TMP triaged helicopter in favour of the scene call. Ten minutes after helicopter diverted towards Algonquin, second scene call came in for a possible trauma case north of Kingston (patient 3). TMP triaged in favour of first scene call until more information was available. Local EMS had picked up patient 2, and helicopter met them on the highway. Paramedics were informed that there was a potential second case, and were told by local EMS that the hospital was 50 minutes away. The team (aircrew and paramedics), assessed the patient and felt that the patient was fit to travel by land EMS. TMP asked if everyone on the team was comfortable with the choice, and agreed with their assessment afterwards. Helicopter then departed for the scene call for patient 3, which was subsequently cancelled.

Major Decisions: 1) Triage between patient 1 and 2. 2) Triage between patient 2 and 3. 3) Whether helicopter should leave patient 2 or patient 3.

Sources of Uncertainty: 1) Patients' conditions (pt. 1, 2, and 3 to varying degrees of certainty).

Case 3 (Role – OM): Shift Change / Resource allocation example.

OM coming on shift came in at 5am, got OCC and paramedic/pilot schedules, incident reports, summary report (i.e., weather and ongoing transfers) from previous OM, and e-mails about the day. Reports stated that there were two OCC staff who had booked the day off in the OCC and that two people were coming in at 10 am. OM decided that the OCC could manage short-staffed until 10am. External schedule stated that Kenora had two paramedics with 6-7 hour late starts due to massive OTs, London only had a single paramedic until noon (with pilots starting at 8am), and Timmins had medics who started at 12:10. OM asked Kenora to do a call-out (i.e., a request for a medic to come in off schedule), but no one responded; base manager filled in as a medic. OM also elected not to do a call-out for London (because it was a quiet morning) or Timmins (since the aircraft had dutied out in Sudbury the shift before). Prior to start of shift, OM debriefed his team about an event that had the RCMP controlling the airspace around Ottawa for the day.

Major Decisions: 1) OCC staffing decision for the day. 2) Possible call-outs for Kenora, Timmins, and London.

Sources of Uncertainty: 1) Call volume for the morning for each of the regions. 2) Ability for Ornge's currently available resources to handle incoming calls (i.e., maintaining coverage).

Case 4 (Role – OM): Finding the most effective resource case.

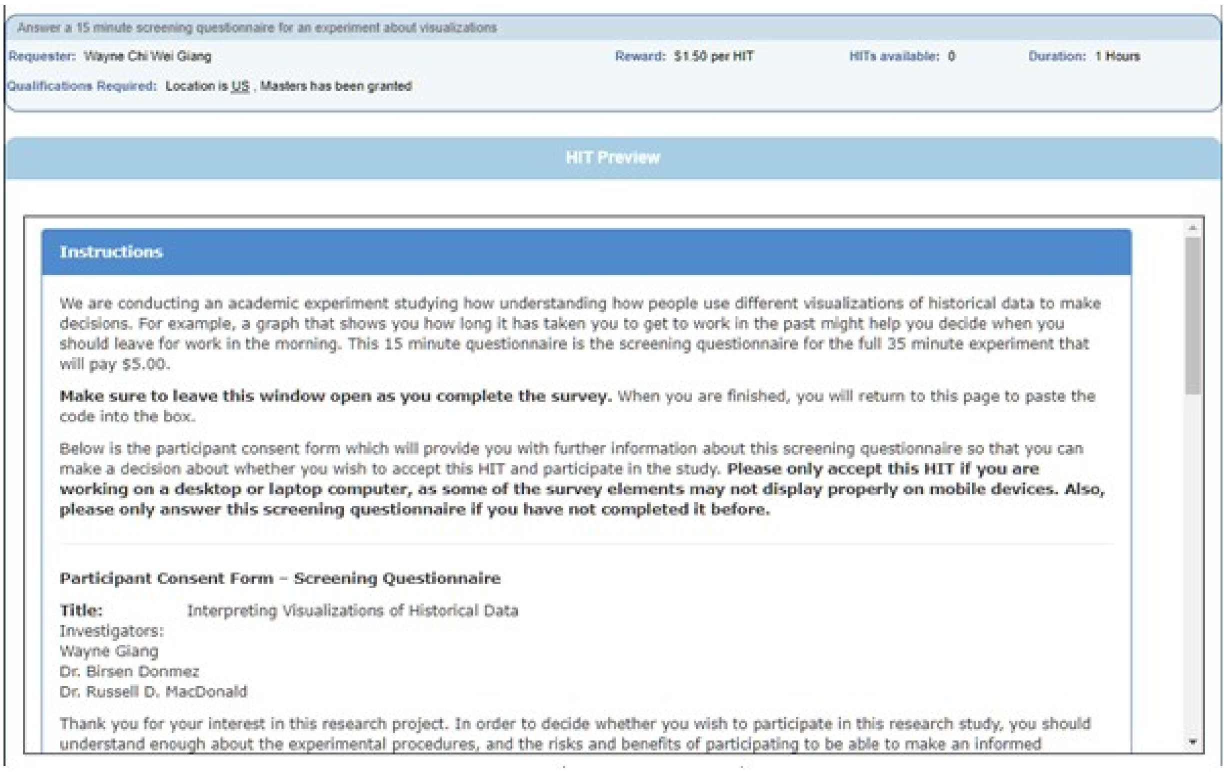
Call comes in from Attawapiskat at 0030 for the transfer of two psych patients who required Ontario Province Police (OPP) escort (2 escorts per patient) to Moose Factory both which are ACP. 793 (Moosonee) helicopter is required for the river hop (note: the Moose Factory hospital is on an island across the river from the Ornge base and patients require either a helicopter trip, a river hop, or a ferry to bring them to the hospital; due to the swollen river levels the ferry was not available as an option), but only has a Primary Care Paramedic (PCP) crew. 790 (Thunder bay fixed-wing) has an Advanced Care Paramedic (ACP) crew but cannot fit both OPP escorts. Planners/OM arrange for 790 to pick up OPP and then meet 793 in Attawapiskat, where the ACP paramedics will transfer onto 793 to transfer the patients. However, call was delayed until next shift due to declined weather checks and OPP not being available until 0800 in Timmins. Next OM starts shift at 0530, and asks a planner to check the availability of an Standing Arrangement (SA) carrier (3rd party medical air transfer services that perform fixed wing patient transfers on a fee-for-service basis) with ACP capabilities (Plan A). While waiting for SA carrier to respond, OM asks planner to plan a transfer where 796 (Timmins fixed-wing with an ACP crew) picks up OPP, and meets 793 in Moosonee where they will transfer ACP paramedics and OPP officers onto 793 to complete the transfers in Attawapiskat (Plan B), with 793 remaining in Moosonee with a temporary PCP crew. SA carrier responds that they have an ACP crew available at 1330, and OM asks TMP whether the patients can wait. The TMP says no, so the OM goes with Plan B. While transfers were being completed, a new emergent transfer (PCP) comes in between Hearst and Timmins. Planner (who was filling in during a break) assigned 796 (with the PCP crew) to the transfer. Once the original planner returns from break, they ask the OM whether another resource (790), which was just finishing up a call could be used instead. The OM agrees, and keeps 796 in Moosonee to await the completion of the Attawapiskat transfers.

Major Decisions: 1) Best way to get an ACP crew, the two OPP escorts, and a patient to Moose Factory. 2) Whether to use a SA carrier or two dedicated Ornge resources. 3) Whether to use 793 or 796 for the Hearst transfer.

Sources of Uncertainty: 1) Availability of ACP SA carrier resource. 2) The opportunity cost of using two dedicated Ornge resources versus the additional financial cost of the SA carrier. 3) The opportunity cost of taking 796 out of Moosonee for a transfer, while its original crew was tied up with the Attawapiskat transfer. 4) The future volume of transfers across the province (i.e., maintaining coverage).

Appendix C – Experimental Materials for Chapter 6

Amazon Mechanical Turk HIT Description for Screening Questionnaire


Answer a 15 minute screening questionnaire for an experiment about visualizations

Requester: Wayne Chi Wei Giang | Reward: $1.50 per HIT | HITs available: 0 | Duration: 1 Hours

Qualifications Required: Location is US , Masters has been granted

HIT Preview

Instructions

We are conducting an academic experiment studying how understanding how people use different visualizations of historical data to make decisions. For example, a graph that shows you how long it has taken you to get to work in the past might help you decide when you should leave for work in the morning. This 15 minute questionnaire is the screening questionnaire for the full 35 minute experiment that will pay $5.00.

Make sure to leave this window open as you complete the survey. When you are finished, you will return to this page to paste the code into the box.

Below is the participant consent form which will provide you with further information about this screening questionnaire so that you can make a decision about whether you wish to accept this HIT and participate in the study. **Please only accept this HIT if you are working on a desktop or laptop computer, as some of the survey elements may not display properly on mobile devices. Also, please only answer this screening questionnaire if you have not completed it before.**

Participant Consent Form – Screening Questionnaire

Title: Interpreting Visualizations of Historical Data
Investigators:
Wayne Giang
Dr. Birsen Donmez
Dr. Russell D. MacDonald

Thank you for your interest in this research project. In order to decide whether you wish to participate in this research study, you should understand enough about the experimental procedures, and the risks and benefits of participating to be able to make an informed

Consent Form for Screening Questionnaire

Visualization_Screening

Participant Consent Form

Participant Consent Form – Screening Questionnaire

Title: Interpreting Visualizations of Historical Data
Investigators:
Wayne Giang
Dr. Birsen Donmez
Dr. Russell D. MacDonald

Thank you for your interest in this research project. In order to decide whether you wish to participate in this research study, you should understand enough about the experimental procedures, and the risks and benefits of participating to be able to make an informed decision. This is known as the informed consent process.

If at any point you feel as though any of the following details are unclear, or if you have any other questions, comments, or concerns, please feel free to contact me using the contact information provided above. **If you decide that you would like to participate, please check off all checkboxes at the bottom of this page. If you do not wish to participate, there is no need to return the form and you may close this window and return the HIT.**

Purpose
We are interested in understanding how people use different visualizations of historical data to make decisions. For example, a graph that shows you how long it has taken you to get to work in the past might help you decide when you should leave for work in the morning.

This current survey is the screening questionnaire for the experiment, and if you qualify you may be invited back to participate in the experiment.

As a participant you will be asked to:

- Complete four questionnaires that relate to personality factors and demographics.
- Read about concepts that are important in understanding visualizations of historical data, and respond to a short 8-question test on these concepts.

Procedure
There are two parts to this survey. In the first part you will be asked to fill out a series of four questionnaires that relate to demographics data and personality factors. The first questionnaire asks for demographics data that may be related to understanding visualizations of historical data such as age, profession, and math education. The second questionnaire examines how likely you are to use numbers in your everyday life. The third and fourth questionnaires examine personality factors that may be related to how you would use data in decision-making. In the second part, you will be presented with information about how historical data is summarized and described. You will be asked to answer a short 8-question test on these concepts.

Risks
There are no major risks involved with this experiment; the tasks are not physiologically demanding, or psychologically stressing.

Benefits
There are several benefits to conducting this study. The most important benefit is your contribution to research in information visualization and decision support systems, which will guide the development of future decision support systems for expert decision makers.

Compensation
This survey should take approximately 15 minutes, and you will be compensated $1.50 upon completion of all the questions. **Based on the results of this survey, you may be invited to participate in a 35 minute experiment, which will be compensated $5.00.**

Confidentiality
All information obtained during the study will be held in strict confidence. You will be identified with a study number only, and this study number will only be identifiable by the researchers. No names or identifying information will be used in any publication or presentation. No information identifying you will be transferred outside the investigators in this study.

Participation
Your participation in this study is voluntary. You can choose to not participate or withdraw at any time by closing this window and returning the HIT before completing all the questions. If you choose to withdraw, you will not be compensated for the study and your data will be deleted.

Questions
If you have any general questions about this study, please call 416.978.0881 or email wayne.giang@mail.utoronto.ca. You can also contact the Office of Research Ethics at ethics.review@utoronto.ca or 416-946-3273, if you have questions about your rights as participants.

1. To be completed by participants, please check all three if you agree to participate in this study: *

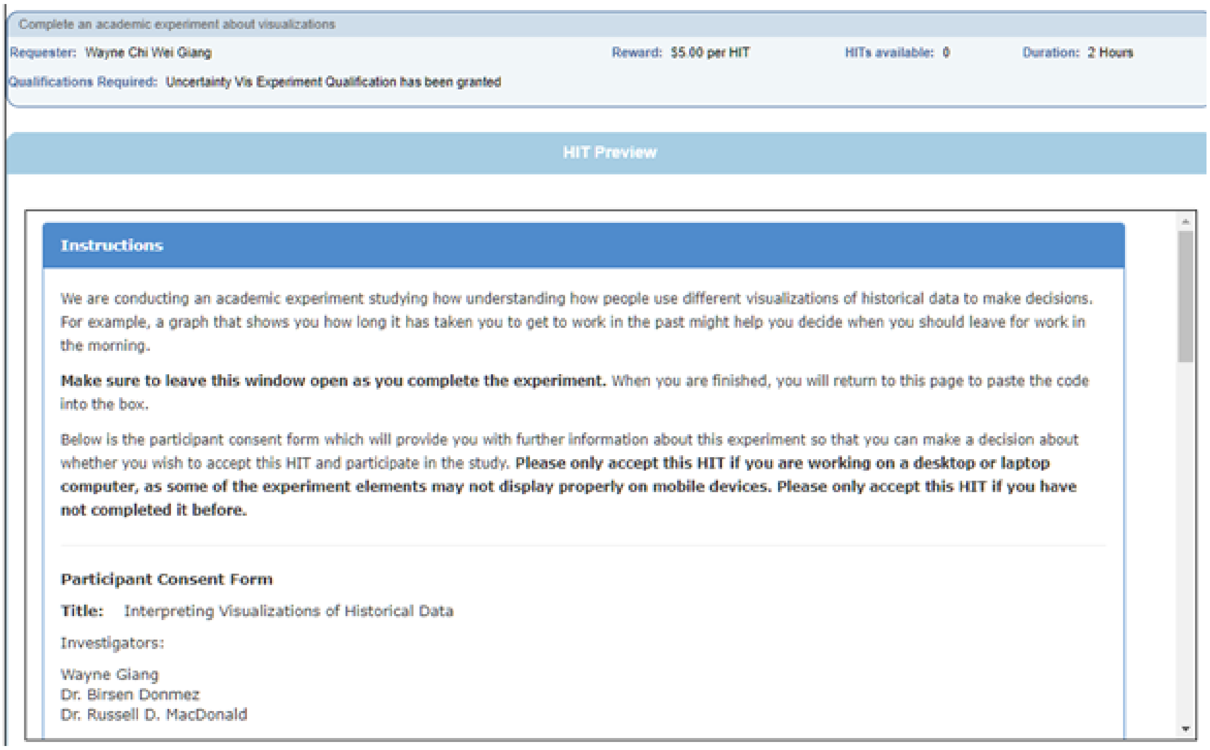
- ☐ I have read this consent form and I understand the research and what is expect of me.
- ☐ I understand that I am free to withdraw before or anytime during the study without the need to give any explanation by closing the browser window.
- ☐ I agree to participate in this study.

If I do not wish to participate in the research, I can just close this Internet browser window without agreeing to the above.

Next

0%

Amazon Mechanical Turk HIT for Experiment



Complete an academic experiment about visualizations

Requester: Wayne Chi Wei Giang | Reward: $5.00 per HIT | HITs available: 0 | Duration: 2 Hours

Qualifications Required: Uncertainty Vis Experiment Qualification has been granted

HIT Preview

Instructions

We are conducting an academic experiment studying how understanding how people use different visualizations of historical data to make decisions. For example, a graph that shows you how long it has taken you to get to work in the past might help you decide when you should leave for work in the morning.

Make sure to leave this window open as you complete the experiment. When you are finished, you will return to this page to paste the code into the box.

Below is the participant consent form which will provide you with further information about this experiment so that you can make a decision about whether you wish to accept this HIT and participate in the study. **Please only accept this HIT if you are working on a desktop or laptop computer, as some of the experiment elements may not display properly on mobile devices. Please only accept this HIT if you have not completed it before.**

Participant Consent Form

Title: Interpreting Visualizations of Historical Data

Investigators:

Wayne Giang
Dr. Birsen Donmez
Dr. Russell D. MacDonald

Consent Form for Experiment

Participant Consent Form

Title: Interpreting Visualizations of Historical Data

Thank you for your interest in this research project. In order to decide whether you wish to participate in this research study, you should understand enough about the experimental procedures, and the risks and benefits of participating to be able to make an informed decision. This is known as the informed consent process.

If at any point you feel that the following details are unclear, or if you have any other questions or concerns, please feel free to contact me using the contact information provided above. If you decide that you would like to participate, please check off all checkboxes at the bottom of the page and enter your MTurk Worker ID. If you do not wish to participate, there is no need to complete the form and you may close this window and return the HIT.

Purpose

We are interested in understanding how people use different visualizations of historical data to make decisions. For example, a graph that shows you how long it has taken you to get to work in the past might help you decide when you should leave for work in the morning.

As a participant you will be asked to:

- Complete an experimental task where you make decisions based on historical data presented using different visualizations.
- Complete a questionnaire about the strategies you used during the experiment.

Procedure

There are two parts to this study. In the first part, you will take the role of a space rover mission commander who is responsible for estimating how long it will take rovers to complete mission-related tasks (e.g., scanning rocks, drilling for samples). You will be shown different visualizations of how long it took each rover to complete the task in the past to help you with your estimates. In the second part of the study, you will fill out a questionnaire about the strategies you used while making these task time estimates.

Risks

There are no major risks involved with this experiment, the tasks are not physiologically demanding, or psychologically stressing.

Benefits

There are several benefits to conducting this study. The most important benefit is your contribution to research in information visualization and decision support systems, which will guide the development of future decision support systems for expert decision makers such as medical dispatchers.

Compensation

This experiment will take approximately 35 minutes. You will receive $5.00 upon completion of the experiment.

Confidentiality

All information obtained during the study will be held in strict confidence. You will be identified with a study number only, and this study number will only be identifiable by the primary investigator. No names or identifying information will be used in any publication or presentation. No information identifying you will be transferred outside the investigators in this study.

Participation

Your participation in this study is voluntary. You can choose to not participate or withdraw at any time by closing this window before completing all the questions and returning the HIT. If you choose to withdraw, your data from this experiment will be deleted. If you choose to withdraw you will not be compensated for this study.

Questions

If you have any general questions about this study, please call 416.978.0881 or email wayne.giang [at] mail.utoronto.ca. You can also contact the Office of Research Ethics at ethics.review [at] utoronto.ca or 416-946-3273, if you have questions about your rights as participants.

To be completed by participants:

☐ I have read this consent form and I understand the research and what is expected of me.
☐ I understand that I am free to withdraw before or anytime during the study without the need to give any explanation by closing the browser window.
☐ I agree to participate in this study.

MTurk WorkerID: ________ * Please enter your MTurk Worker ID.

Submit

Datasets used for experiment in Chapter 6

Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6
20	10	25	8	6	15
21	11	27	12	8	21
22	12	27	12	8	23
23	12	28	13	9	23
23	12	29	13	9	24
23	13	29	14	10	25
23	13	29	14	10	25
24	13	29	14	11	26
24	14	29	15	11	26
24	14	29	15	11	26
24	14	30	15	12	26
24	14	30	17	12	27
24	14	30	17	12	27
25	14	30	17	13	27
25	14	30	17	13	28
25	14	30	17	13	29
25	14	30	17	13	29
25	14	30	18	13	29
25	14	31	19	14	30
25	15	31	19	14	30
25	15	31	19	15	30
25	15	31	19	15	31
25	15	31	19	16	32
25	15	31	19	16	32
26	15	31	19	16	32
26	15	31	19	17	32
26	15	31	20	17	32
26	16	31	20	18	32
26	16	31	20	18	32
26	16	31	20	19	33
26	16	32	20	19	33
26	16	32	22	19	34
26	16	32	22	20	34
27	16	32	22	21	34
27	16	32	23	21	35
27	16	32	23	21	35
27	17	32	23	21	36
27	17	32	23	21	36
27	17	32	23	21	36
28	17	33	23	22	37
28	17	33	25	22	37

28	17	33	25	23	38
28	17	33	25	24	38
28	17	33	27	25	38
29	18	34	27	25	38
29	19	34	28	25	38
29	19	34	28	27	40
29	20	34	29	29	41
30	20	34	30	31	42
30	20	35	32	34	47

Appendix D – Statistical Models for Chapter 6

Predictions on the Salient Central Tendency Point

```
proc genmod data = Exp2.data descending;
CLASS ParticipantID PartCond Var scenario_sd;
model deviation = PartCond*Var*scenario_sd/ link = logit dist = binomial type3 wald;
repeated subject = ParticipantID(PartCond);

estimate 'Average Median-only' intercept 1 PartCond*Var*scenario_sd  0.166666667      0.166666667      0        0
        0.166666667     0.166666667     0       0       0.166666667     0.166666667     0       0 /exp;
estimate 'Large SD: Average Median-only' intercept 1 PartCond*Var*scenario_sd 0.333333333      0        0        0
        0.333333333     0       0       0       0.333333333     0       0       0 /exp;
estimate 'Large SD: Mean&SD' intercept 1 PartCond*Var*scenario_sd 0 0      0       0       0       0       0
        0       0       0       1       0       /exp;
estimate 'Large SD: Boxplot' intercept 1 PartCond*Var*scenario_sd 0 0      0       0       0       0       1
        0       0       0       0       0       /exp;
estimate 'Large SD: Dotplot' intercept 1 PartCond*Var*scenario_sd 0 0      1       0       0       0       0
        0       0       0       0       0       /exp;

estimate 'Small SD: Average Median-only' intercept 1 PartCond*Var*scenario_sd 0 0.333333333     0       0       0
        0.333333333     0       0       0       0.333333333     0       0       /exp;
estimate 'Small SD: Mean&SD' intercept 1 PartCond*Var*scenario_sd 0 0      0       0       0       0       0
        0       0       0       0       1       /exp;
estimate 'Small SD: Boxplot' intercept 1 PartCond*Var*scenario_sd 0 0      0       0       0       0       0
        1       0       0       0       0       /exp;
estimate 'Small SD: Dotplot' intercept 1 PartCond*Var*scenario_sd 0 0      0       1       0       0       0
        0       0       0       0       0       /exp;

estimate 'Large vs. Small SD' PartCond*Var*scenario_sd 0.166666667 -0.166666667      0.166666667 -0.166666667
0.166666667 -0.166666667 0.166666667 -0.166666667 0.166666667 -0.166666667 0.166666667   -0.166666667 /exp;

estimate 'Dotplot vs Boxplot' PartCond*Var*scenario_sd 0  0     0.5     0.5     0       0       -0.5    -0.5
        0       0       0       0 /exp;
estimate 'Dotplot vs Mean&SD' PartCond*Var*scenario_sd 0       0       0.5     0.5     0       0       0
        0       0       0       -0.5    -0.5 /exp;
estimate 'Dotplot vs Median-only' PartCond*Var*scenario_sd -0.166666667      -0.166666667     0.5     0.5     -
0.166666667     -0.166666667    0       0       -0.166666667    -0.166666667    0       0 /exp;
estimate 'Boxplot vs Mean&SD' PartCond*Var*scenario_sd 0       0       0       0       0       0       0.5
        0.5     0       0       -0.5    -0.5 /exp;
estimate 'Boxplot vs Median-only' PartCond*Var*scenario_sd -0.166666667      -0.166666667     0       0       -
0.166666667     -0.166666667    0.5     0.5     -0.166666667    -0.166666667    0       0 /exp;
estimate 'Mean&SD vs Median-only' PartCond*Var*scenario_sd  -0.166666667             -0.166666667     0       0
        -0.166666667    -0.166666667    0       0       -0.166666667    -0.166666667    0.5     0.5 /exp;

estimate 'Dotplot: Large vs Small SD' PartCond*Var*scenario_sd 0  0     1       -1      0       0       0
        0       0       0       0       0 / exp;
estimate 'Boxplot: Large vs Small SD' PartCond*Var*scenario_sd 0  0     0       0       0       0       1
        -1      0       0       0       0 / exp;
estimate 'Mean&SD: Large vs Small SD' PartCond*Var*scenario_sd 0       0       0       0       0       0
        0       0       0       0       1       -1 /exp;
estiamte 'Median : Large vs Small SD' PartCond*Var*scenario_sd 0.333333333           -0.333333333     0       0
        0.333333333     -0.333333333    0       0       0.333333333     -0.333333333    0       0 / exp;

estimate 'Large SD: Dot vs Box' PartCond*Var*scenario_sd 0     0       1       0       0       0       -1
        0       0       0       0       0 / exp;
estimate 'Large SD: Dot vs SD ' PartCond*Var*scenario_sd 0     0       1       0       0       0       0
        0       0       0       -1      0 / exp;
estimate 'Large SD: Dot vs Med' PartCond*Var*scenario_sd -0.333333333   0       1       0       -0.333333333
        0       0       0       -0.333333333    0       0       0 /exp;
estimate 'Large SD: Box vs SD ' PartCond*Var*scenario_sd 0     0       0       0       0       0       1
        0       0       0       -1      0 /exp;
```

```
estimate 'Large SD: Box vs Med' PartCond*Var*scenario_sd -0.333333333     0     0     0     -0.333333333
     0     1     0     -0.333333333     0     0     0 /exp;
estimate 'Large SD: SD  vs Med' PartCond*Var*scenario_sd -0.333333333     0     0     0     -0.333333333
     0     0     0     -0.333333333     0     1     0 /exp;

estimate 'Small SD: Dot vs Box' PartCond*Var*scenario_sd 0 0     0     1     0     0     0     -1
     0     0     0     0     / exp;
estimate 'Small SD: Dot vs SD ' PartCond*Var*scenario_sd 0 0     0     1     0     0     0     0
     0     0     0     -1     / exp;
estimate 'Small SD: Dot vs Med' PartCond*Var*scenario_sd 0 -0.333333333     0     1     0     -0.333333333
     0     0     0     -0.333333333     0     0     /exp;
estimate 'Small SD: Box vs SD ' PartCond*Var*scenario_sd 0 0     0     0     0     0     0     1
     0     0     0     -1     /exp;
estimate 'Small SD: Box vs Med' PartCond*Var*scenario_sd 0 -0.333333333     0     0     0     -0.333333333
     0     1     0     -0.333333333     0     0     /exp;
estimate 'Small SD: SD  vs Med' PartCond*Var*scenario_sd 0 -0.333333333     0     0     0     -0.333333333
     0     0     0     -0.333333333     0     1     /exp;
run; quit;
```

Analysis Of GEE Parameter Estimates

Empirical Standard Error Estimates

Parameter				Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept				0.7691	0.4553	-0.1233	1.6616	1.69	0.0912
PartCon*Var*scenario	Dotplot	FALSE	2	-0.9698	0.5881	-2.1225	0.1829	-1.65	0.0991
PartCon*Var*scenario	Dotplot	FALSE	5	-0.7691	0.6097	-1.9642	0.4259	-1.26	0.2072
PartCon*Var*scenario	Dotplot	TRUE	2	0.6172	0.5697	-0.4994	1.7337	1.08	0.2786
PartCon*Var*scenario	Dotplot	TRUE	5	0.4205	0.6044	-0.7642	1.6051	0.70	0.4867
PartCon*Var*scenario	Boxplot	FALSE	2	-0.3637	0.6141	-1.5672	0.8399	-0.59	0.5537
PartCon*Var*scenario	Boxplot	FALSE	5	-0.1501	0.6535	-1.4310	1.1308	-0.23	0.8184
PartCon*Var*scenario	Boxplot	TRUE	2	0.0782	0.5977	-1.0933	1.2496	0.13	0.8960
PartCon*Var*scenario	Boxplot	TRUE	5	0.3295	0.6531	-0.9507	1.6096	0.50	0.6139
PartCon*Var*scenario	Standard_Deviation	FALSE	2	-0.5685	0.2901	-1.1370	0.0001	-1.96	0.0500
PartCon*Var*scenario	Standard_Deviation	FALSE	5	-0.5009	0.2530	-0.9968	-0.0049	-1.98	0.0478
PartCon*Var*scenario	Standard_Deviation	TRUE	2	-0.3637	0.1747	-0.7061	-0.0212	-2.08	0.0374
PartCon*Var*scenario	Standard_Deviation	TRUE	5	0.0000	0.0000	0.0000	0.0000	.	.

Wald Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
PartCon*Var*scenario	11	29.67	0.0018

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Average Median-only	0.5537	0.4380	0.6638	0.2155	0.2371	0.05	-0.2492	0.6802	0.83	0.3635
Exp(Average Median-only)				1.2404	0.2941	0.05	0.7794	1.9742		
Large SD: Average Median-only	0.5337	0.4179	0.6461	0.1352	0.2381	0.05	-0.3316	0.6019	0.32	0.5703
Exp(Large SD: Average Median-only)				1.1447	0.2726	0.05	0.7178	1.8255		
Large SD: Mean&SD	0.6000	0.3901	0.7786	0.4055	0.4348	0.05	-0.4467	1.2576	0.87	0.3511
Exp(Large SD: Mean&SD)				1.5000	0.6522	0.05	0.6397	3.5171		
Large SD: Boxplot	0.7000	0.5221	0.8329	0.8473	0.3872	0.05	0.0884	1.6062	4.79	0.0286
Exp(Large SD: Boxplot)				2.3333	0.9034	0.05	1.0925	4.9837		
Large SD: Dotplot	0.8000	0.6716	0.8867	1.3863	0.3423	0.05	0.7153	2.0572	16.40	<.0001
Exp(Large SD: Dotplot)				4.0000	1.3693	0.05	2.0449	7.8244		
Small SD: Average Median-only	0.5734	0.4499	0.6884	0.2958	0.2536	0.05	-0.2012	0.7927	1.36	0.2434

Contrast Estimate Results

Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Exp(Small SD: Average Median-only)				1.3442	0.3408	0.05	0.8177	2.2094		
Small SD: Mean&SD	0.6833	0.4692	0.8405	0.7691	0.4553	0.05	-0.1233	1.6616	2.85	0.0912
Exp(Small SD: Mean&SD)				2.1579	0.9826	0.05	0.8840	5.2677		
Small SD: Boxplot	0.7500	0.5451	0.8825	1.0986	0.4683	0.05	0.1809	2.0164	5.50	0.0190
Exp(Small SD: Boxplot)				3.0000	1.4048	0.05	1.1982	7.5110		
Small SD: Dotplot	0.7667	0.6012	0.8775	1.1896	0.3975	0.05	0.4106	1.9686	8.96	0.0028
Exp(Small SD: Dotplot)				3.2857	1.3060	0.05	1.5077	7.1607		
Large vs. Small SD	0.4626	0.4145	0.5113	-0.1500	0.0997	0.05	-0.3454	0.0454	2.26	0.1324
Exp(Large vs. Small SD)				0.8607	0.0858	0.05	0.7079	1.0464		
Dotplot vs Boxplot	0.5781	0.3241	0.7966	0.3150	0.5358	0.05	-0.7351	1.3650	0.35	0.5566
Exp(Dotplot vs Boxplot)				1.3702	0.7341	0.05	0.4795	3.9159		
Dotplot vs Mean&SD	0.6683	0.4053	0.8563	0.7006	0.5531	0.05	-0.3834	1.7847	1.60	0.2053
Exp(Dotplot vs Mean&SD)				2.0150	1.1145	0.05	0.6815	5.9579		
Dotplot vs Median-only	0.7451	0.5790	0.8613	1.0725	0.3845	0.05	0.3188	1.8262	7.78	0.0053

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Exp(Dotplot vs Median-only)				2.9226	1.1239	0.05	1.3755	6.2101		
Boxplot vs Mean&SD	0.5952	0.3113	0.8271	0.3857	0.6018	0.05	-0.7939	1.5652	0.41	0.5217
Exp(Boxplot vs Mean&SD)				1.4706	0.8851	0.05	0.4521	4.7839		
Boxplot vs Median-only	0.6808	0.5107	0.8134	0.7575	0.3646	0.05	0.0429	1.4721	4.32	0.0377
Exp(Boxplot vs Median-only)				2.1329	0.7777	0.05	1.0438	4.3584		
Mean&SD vs Median-only	0.5919	0.4104	0.7514	0.3718	0.3746	0.05	-0.3625	1.1061	0.99	0.3210
Exp(Mean&SD vs Median-only)				1.4504	0.5434	0.05	0.6960	3.0226		
Dotplot: Large vs Small SD	0.5490	0.4043	0.6859	0.1967	0.2980	0.05	-0.3875	0.7809	0.44	0.5093
Exp(Dotplot: Large vs Small SD)				1.2174	0.3628	0.05	0.6788	2.1834		
Boxplot: Large vs Small SD	0.4375	0.3325	0.5484	-0.2513	0.2274	0.05	-0.6970	0.1944	1.22	0.2691
Exp(Boxplot: Large vs Small SD)				0.7778	0.1769	0.05	0.4981	1.2146		
Mean&SD: Large vs Small SD	0.4101	0.3305	0.4947	-0.3637	0.1747	0.05	-0.7061	-0.0212	4.33	0.0374
Exp(Mean&SD: Large vs Small SD)				0.6951	0.1214	0.05	0.4936	0.9790		
Median : Large vs Small SD	0.4599	0.3972	0.5240	-0.1606	0.1309	0.05	-0.4172	0.0959	1.51	0.2198

Contrast Estimate Results										
		Mean					L'Beta			
Label	Mean Estimate	Confidence Limits		L'Beta Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
Exp(Median : Large vs Small SD)				0.8516	0.1115	0.05	0.6589	1.1007		
Large SD: Dot vs Box	0.6316	0.3837	0.8252	0.5390	0.5168	0.05	-0.4739	1.5519	1.09	0.2970
Exp(Large SD: Dot vs Box)				1.7143	0.8860	0.05	0.6225	4.7206		
Large SD: Dot vs SD	0.7273	0.4741	0.8875	0.9808	0.5534	0.05	-0.1038	2.0654	3.14	0.0763
Exp(Large SD: Dot vs SD)				2.6667	1.4757	0.05	0.9014	7.8887		
Large SD: Dot vs Med	0.7775	0.6265	0.8792	1.2511	0.3745	0.05	0.5172	1.9851	11.16	0.0008
Exp(Large SD: Dot vs Med)				3.4943	1.3085	0.05	1.6774	7.2795		
Large SD: Box vs SD	0.6087	0.3320	0.8296	0.4418	0.5822	0.05	-0.6993	1.5829	0.58	0.4479
Exp(Large SD: Box vs SD)				1.5556	0.9056	0.05	0.4970	4.8692		
Large SD: Box vs Med	0.6709	0.5053	0.8027	0.7121	0.3526	0.05	0.0211	1.4032	4.08	0.0434
Exp(Large SD: Box vs Med)				2.0384	0.7187	0.05	1.0213	4.0681		
Large SD: SD vs Med	0.5672	0.3856	0.7323	0.2703	0.3755	0.05	-0.4656	1.0063	0.52	0.4716
Exp(Large SD: SD vs Med)				1.3104	0.4920	0.05	0.6277	2.7353		
Small SD: Dot vs Box	0.5227	0.2473	0.7850	0.0910	0.6142	0.05	-1.1128	1.2948	0.02	0.8823

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Exp(Small SD: Dot vs Box)				1.0952	0.6727	0.05	0.3286	3.6502		
Small SD: Dot vs SD	0.6036	0.3177	0.8327	0.4205	0.6044	0.05	-0.7642	1.6051	0.48	0.4867
Exp(Small SD: Dot vs SD)				1.5226	0.9203	0.05	0.4657	4.9783		
Small SD: Dot vs Med	0.7097	0.4968	0.8582	0.8938	0.4626	0.05	-0.0128	1.8005	3.73	0.0533
Exp(Small SD: Dot vs Med)				2.4444	1.1308	0.05	0.9872	6.0525		
Small SD: Box vs SD	0.5816	0.2788	0.8334	0.3295	0.6531	0.05	-0.9507	1.6096	0.25	0.6139
Exp(Small SD: Box vs SD)				1.3902	0.9080	0.05	0.3865	5.0009		
Small SD: Box vs Med	0.6906	0.4989	0.8334	0.8028	0.4118	0.05	-0.0043	1.6100	3.80	0.0512
Exp(Small SD: Box vs Med)				2.2319	0.9191	0.05	0.9957	5.0027		
Small SD: SD vs Med	0.6162	0.4235	0.7782	0.4734	0.3988	0.05	-0.3083	1.2550	1.41	0.2353
Exp(Small SD: SD vs Med)				1.6054	0.6403	0.05	0.7347	3.5080		

Direction of Predictions relative to Central Tendency Point

```
proc genmod data = Exp2.dev descending;
CLASS ParticipantID PartCond Var scenario_sd;
model AboveCT = Var*PartCond*scenario_sd/ link = logit dist = binomial type3 wald;
repeated subject = ParticipantID(PartCond);

estimate 'Large vs. Small SD' PartCond*Var*scenario_sd 0.166666667 -0.166666667     0.166666667 -0.166666667
0.166666667 -0.166666667 0.166666667 -0.166666667 0.166666667 -0.166666667 0.166666667   -0.166666667 /exp;
```

```
estimate 'Dotplot vs Boxplot' PartCond*Var*scenario_sd 0  0          0.5         0.5         0           0           -0.5        -0.5
            0           0           0           0 /exp;
estimate 'Dotplot vs Mean&SD' PartCond*Var*scenario_sd 0          0           0.5         0.5         0           0           0
            0           0           0           -0.5        -0.5 /exp;
estimate 'Dotplot vs Median-only' PartCond*Var*scenario_sd -0.166666667     -0.166666667      0.5         0.5         -
0.166666667       -0.166666667      0           0           -0.166666667      -0.166666667      0           0 /exp;
estimate 'Boxplot vs Mean&SD' PartCond*Var*scenario_sd 0          0           0           0           0           0           0.5
            0.5         0           0           -0.5        -0.5 /exp;
estimate 'Boxplot vs Median-only' PartCond*Var*scenario_sd -0.166666667     -0.166666667      0           0           -
0.166666667       -0.166666667      0.5         0.5         -0.166666667      -0.166666667      0           0 /exp;
estimate 'Mean&SD vs Median-only' PartCond*Var*scenario_sd  -0.166666667           -0.166666667      0           0
            -0.166666667      -0.166666667      0           0           -0.166666667      -0.166666667      0.5         0.5 /exp;

estimate 'Dotplot: Large vs Small SD' PartCond*Var*scenario_sd 0  0           1           -1          0           0           0
            0           0           0           0           0 / exp;
estimate 'Boxplot: Large vs Small SD' PartCond*Var*scenario_sd 0  0           0           0           0           0           1
            -1          0           0           0           0 / exp;
estimate 'Mean&SD: Large vs Small SD' PartCond*Var*scenario_sd 0            0           0           0           0           0
            0           0           0           0           1           -1 /exp;
estiamte 'Median : Large vs Small SD' PartCond*Var*scenario_sd 0.333333333            -0.333333333      0           0
            0.333333333       -0.333333333      0           0           0.333333333       -0.333333333      0           0 / exp;

estimate 'Large SD: Dot vs Box' PartCond*Var*scenario_sd 0          0           1           0           0           0           -1
            0           0           0           0           0 / exp;
estimate 'Large SD: Dot vs SD ' PartCond*Var*scenario_sd 0          0           1           0           0           0           0
            0           0           0           -1          0 / exp;
estimate 'Large SD: Dot vs Med' PartCond*Var*scenario_sd -0.333333333     0           1           0           -0.333333333
            0           0           0           -0.333333333      0           0           0 /exp;
estimate 'Large SD: Box vs SD ' PartCond*Var*scenario_sd 0          0           0           0           0           0           1
            0           0           0           -1          0 /exp;
estimate 'Large SD: Box vs Med' PartCond*Var*scenario_sd -0.333333333     0           0           0           -0.333333333
            0           1           0           -0.333333333      0           0           0 /exp;
estimate 'Large SD: SD  vs Med' PartCond*Var*scenario_sd -0.333333333     0           0           0           -0.333333333
            0           0           0           -0.333333333      0           1           0 /exp;

estimate 'Small SD: Dot vs Box' PartCond*Var*scenario_sd 0 0        0           1           0           0           0           -1
            0           0           0           0           / exp;
estimate 'Small SD: Dot vs SD ' PartCond*Var*scenario_sd 0 0        0           1           0           0           0           0
            0           0           0           -1          / exp;
estimate 'Small SD: Dot vs Med' PartCond*Var*scenario_sd 0 -0.333333333     0           1           0           -0.333333333
            0           0           0           -0.333333333      0           0           /exp;
estimate 'Small SD: Box vs SD ' PartCond*Var*scenario_sd 0 0        0           0           0           0           0           1
            0           0           0           -1          /exp;
estimate 'Small SD: Box vs Med' PartCond*Var*scenario_sd 0 -0.333333333     0           0           0           -0.333333333
            0           1           0           -0.333333333      0           0           /exp;
estimate 'Small SD: SD  vs Med' PartCond*Var*scenario_sd 0 -0.333333333     0           0           0           -0.333333333
            0           0           0           -0.333333333      0           1           /exp;
run; quit;
```

Analysis Of GEE Parameter Estimates									
Empirical Standard Error Estimates									
Parameter				Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept				1.5805	0.3965	0.8033	2.3576	3.99	<.0001
PartCon*Var*scenario	Dotplot	FALSE	2	-0.0988	0.5928	-1.2608	1.0631	-0.17	0.8676

Analysis Of GEE Parameter Estimates									
Empirical Standard Error Estimates									
Parameter				Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
PartCon*Var*scenario	Dotplot	FALSE	5	-0.5688	0.5647	-1.6757	0.5380	-1.01	0.3138
PartCon*Var*scenario	Dotplot	TRUE	2	2.2697	1.0818	0.1493	4.3900	2.10	0.0359
PartCon*Var*scenario	Dotplot	TRUE	5	-0.9518	0.4525	-1.8387	-0.0649	-2.10	0.0354
PartCon*Var*scenario	Boxplot	FALSE	2	-0.4818	0.6270	-1.7108	0.7471	-0.77	0.4422
PartCon*Var*scenario	Boxplot	FALSE	5	-0.5157	0.5545	-1.6026	0.5711	-0.93	0.3523
PartCon*Var*scenario	Boxplot	TRUE	2	-0.8873	0.5041	-1.8754	0.1008	-1.76	0.0784
PartCon*Var*scenario	Boxplot	TRUE	5	-0.4520	0.4645	-1.3623	0.4584	-0.97	0.3305
PartCon*Var*scenario	Standard_Deviation	FALSE	2	-0.5996	0.2881	-1.1644	-0.0349	-2.08	0.0374
PartCon*Var*scenario	Standard_Deviation	FALSE	5	-0.7050	0.4148	-1.5179	0.1080	-1.70	0.0892
PartCon*Var*scenario	Standard_Deviation	TRUE	2	-0.8873	0.3266	-1.5275	-0.2471	-2.72	0.0066
PartCon*Var*scenario	Standard_Deviation	TRUE	5	0.0000	0.0000	0.0000	0.0000	.	.

Wald Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
PartCon*Var*scenario	11	29.30	0.0020

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Large vs. Small SD	0.6030	0.4929	0.7036	0.4180	0.2278	0.05	-0.0285	0.8646	3.37	0.0665
Exp(Large vs. Small SD)				1.5190	0.3461	0.05	0.9719	2.3740		

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Dotplot vs Boxplot	0.7906	0.5492	0.9213	1.3286	0.5771	0.05	0.1975	2.4596	5.30	0.0213
Exp(Dotplot vs Boxplot)				3.7756	2.1789	0.05	1.2183	11.7007		
Dotplot vs Mean&SD	0.7507	0.4541	0.9160	1.1026	0.6565	0.05	-0.1842	2.3893	2.82	0.0931
Exp(Dotplot vs Mean&SD)				3.0119	1.9774	0.05	0.8318	10.9061		
Dotplot vs Median-only	0.7602	0.5094	0.9064	1.1539	0.5696	0.05	0.0375	2.2703	4.10	0.0428
Exp(Dotplot vs Median-only)				3.1706	1.8059	0.05	1.0383	9.6821		
Boxplot vs Mean&SD	0.4437	0.2383	0.6704	-0.2260	0.4776	0.05	-1.1621	0.7101	0.22	0.6361
Exp(Boxplot vs Mean&SD)				0.7977	0.3810	0.05	0.3128	2.0343		
Boxplot vs Median-only	0.4564	0.3395	0.5784	-0.1747	0.2505	0.05	-0.6656	0.3162	0.49	0.4856
Exp(Boxplot vs Median-only)				0.8397	0.2103	0.05	0.5140	1.3719		
Mean&SD vs Median-only	0.5128	0.3423	0.6804	0.0513	0.3593	0.05	-0.6529	0.7555	0.02	0.8864
Exp(Mean&SD vs Median-only)				1.0527	0.3782	0.05	0.5205	2.1288		
Dotplot: Large vs Small SD	0.9616	0.7709	0.9947	3.2215	1.0244	0.05	1.2137	5.2294	9.89	0.0017
Exp(Dotplot: Large vs Small SD)				25.0667	25.6795	0.05	3.3658	186.6843		

Contrast Estimate Results											
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Boxplot: Large vs Small SD	0.3929	0.2940	0.5014	-0.4353	0.2249	0.05	-0.8762	0.0056	3.75	0.0530
Exp(Boxplot: Large vs Small SD)				0.6471	0.1455	0.05	0.4164	1.0056		
Mean&SD: Large vs Small SD	0.2917	0.1784	0.4385	-0.8873	0.3266	0.05	-1.5275	-0.2471	7.38	0.0066
Exp(Mean&SD : Large vs Small SD)				0.4118	0.1345	0.05	0.2171	0.7811		
Median : Large vs Small SD	0.5506	0.4197	0.6748	0.2031	0.2689	0.05	-0.3240	0.7302	0.57	0.4501
Exp(Median : Large vs Small SD)				1.2252	0.3295	0.05	0.7232	2.0754		
Large SD: Dot vs Box	0.9592	0.7488	0.9946	3.1570	1.0536	0.05	1.0920	5.2220	8.98	0.0027
Exp(Large SD: Dot vs Box)				23.5000	24.7595	0.05	2.9802	185.3068		
Large SD: Dot vs SD	0.9592	0.7268	0.9952	3.1570	1.1115	0.05	0.9786	5.3354	8.07	0.0045
Exp(Large SD: Dot vs SD)				23.5000	26.1194	0.05	2.6607	207.5616		
Large SD: Dot vs Med	0.9348	0.6470	0.9912	2.6631	1.0497	0.05	0.6058	4.7205	6.44	0.0112
Exp(Large SD: Dot vs Med)				14.3411	15.0537	0.05	1.8327	112.2226		
Large SD: Box vs SD	0.5000	0.2484	0.7516	-0.0000	0.5649	0.05	-1.1073	1.1073	0.00	1.0000
Exp(Large SD: Box vs SD)				1.0000	0.5649	0.05	0.3305	3.0261		

Contrast Estimate Results									
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Large SD: Box vs Med	0.3790	0.2491	0.5289	-0.4939	0.3110	0.05	-1.1035	0.1157	2.52	0.1123
Exp(Large SD: Box vs Med)				0.6103	0.1898	0.05	0.3317	1.1227		
Large SD: SD vs Med	0.3790	0.1973	0.6024	-0.4939	0.4639	0.05	-1.4031	0.4153	1.13	0.2870
Exp(Large SD: SD vs Med)				0.6103	0.2831	0.05	0.2458	1.5149		
Small SD: Dot vs Box	0.3776	0.2427	0.5345	-0.4999	0.3256	0.05	-1.1381	0.1384	2.36	0.1248
Exp(Small SD: Dot vs Box)				0.6066	0.1975	0.05	0.3204	1.1484		
Small SD: Dot vs SD	0.2785	0.1372	0.4838	-0.9518	0.4525	0.05	-1.8387	-0.0649	4.42	0.0354
Exp(Small SD: Dot vs SD)				0.3860	0.1747	0.05	0.1590	0.9371		
Small SD: Dot vs Med	0.4121	0.2784	0.5601	-0.3553	0.3046	0.05	-0.9524	0.2418	1.36	0.2435
Exp(Small SD: Dot vs Med)				0.7010	0.2135	0.05	0.3858	1.2735		
Small SD: Box vs SD	0.3889	0.2039	0.6126	-0.4520	0.4645	0.05	-1.3623	0.4584	0.95	0.3305
Exp(Small SD: Box vs SD)				0.6364	0.2956	0.05	0.2561	1.5815		
Small SD: Box vs Med	0.5361	0.3952	0.6714	0.1445	0.2908	0.05	-0.4254	0.7145	0.25	0.6191
Exp(Small SD: Box vs Med)				1.1555	0.3360	0.05	0.6535	2.0431		
Small SD: SD vs Med	0.6449	0.4586	0.7956	0.5965	0.3891	0.05	-0.1662	1.3592	2.35	0.1253

Contrast Estimate Results										
		Mean					L'Beta			
Label	Mean Estimate	Confidence Limits		L'Beta Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
Exp(Small SD: SD vs Med)				1.8158	0.7066	0.05	0.8469	3.8931		

Distance between Prediction and Central Tendency Point

```
proc genmod data = Exp2.data_dev descending;
CLASS ParticipantID PartCond Var scenario_sd;
model Dev_Cut = Var*PartCond*scenario_sd/ link = clogit dist = mult type3 wald ;
repeated subject = ParticipantID(PartCond);

estimate 'Large vs. Small SD' PartCond*Var*scenario_sd 0.166666667 -0.166666667 0.166666667 -0.166666667
0.166666667 -0.166666667 0.166666667 -0.166666667 0.166666667 -0.166666667 0.166666667 -0.166666667 /exp;

estimate 'Dotplot vs Boxplot' PartCond*Var*scenario_sd 0 0 0.5 0.5 0 0 -0.5 -0.5
 0 0 0 0 /exp;
estimate 'Dotplot vs Mean&SD' PartCond*Var*scenario_sd 0 0 0.5 0.5 0 0 0
 0 0 0 -0.5 -0.5 /exp;
estimate 'Dotplot vs Median-only' PartCond*Var*scenario_sd -0.166666667 -0.166666667 0.5 0.5 -
0.166666667 -0.166666667 0 0 -0.166666667 -0.166666667 0 0 /exp;
estimate 'Boxplot vs Mean&SD' PartCond*Var*scenario_sd 0 0 0 0 0 0 0.5
 0.5 0 0 -0.5 -0.5 /exp;
estimate 'Boxplot vs Median-only' PartCond*Var*scenario_sd -0.166666667 -0.166666667 0 0 -
0.166666667 -0.166666667 0.5 0.5 -0.166666667 -0.166666667 0 0 /exp;
estimate 'Mean&SD vs Median-only' PartCond*Var*scenario_sd -0.166666667 -0.166666667 0 0
 -0.166666667 -0.166666667 0 0 -0.166666667 -0.166666667 0.5 0.5 /exp;

estimate 'Dotplot: Large vs Small SD' PartCond*Var*scenario_sd 0 0 1 -1 0 0 0
 0 0 0 0 0 / exp;
estimate 'Boxplot: Large vs Small SD' PartCond*Var*scenario_sd 0 0 0 0 0 0 1
 -1 0 0 0 0 / exp;
estimate 'Mean&SD: Large vs Small SD' PartCond*Var*scenario_sd 0 0 0 0 0 0
 0 0 0 0 1 -1 /exp;
estiamte 'Median : Large vs Small SD' PartCond*Var*scenario_sd 0.333333333 -0.333333333 0 0
 0.333333333 -0.333333333 0 0 0.333333333 -0.333333333 0 0 / exp;

estimate 'Large SD: Dot vs Box' PartCond*Var*scenario_sd 0 0 1 0 0 0 -1
 0 0 0 0 0 / exp;
estimate 'Large SD: Dot vs SD ' PartCond*Var*scenario_sd 0 0 1 0 0 0 0
 0 0 0 -1 0 / exp;
estimate 'Large SD: Dot vs Med' PartCond*Var*scenario_sd -0.333333333 0 1 0 -0.333333333
 0 0 0 -0.333333333 0 0 0 /exp;
estimate 'Large SD: Box vs SD ' PartCond*Var*scenario_sd 0 0 0 0 0 0 1
 0 0 0 -1 0 /exp;
estimate 'Large SD: Box vs Med' PartCond*Var*scenario_sd -0.333333333 0 0 0 -0.333333333
 0 1 0 -0.333333333 0 0 0 /exp;
estimate 'Large SD: SD  vs Med' PartCond*Var*scenario_sd -0.333333333 0 0 0 -0.333333333
 0 0 0 -0.333333333 0 1 0 /exp;

estimate 'Small SD: Dot vs Box' PartCond*Var*scenario_sd 0 0 0 1 0 0 0 -1
 0 0 0 0 / exp;
estimate 'Small SD: Dot vs SD ' PartCond*Var*scenario_sd 0 0 0 1 0 0 0 0
 0 0 0 -1 / exp;
```

```
estimate 'Small SD: Dot vs Med' PartCond*Var*scenario_sd 0 -0.333333333    0    1    0    -0.333333333
    0    0    0    -0.333333333    0    0    /exp;
estimate 'Small SD: Box vs SD ' PartCond*Var*scenario_sd 0 0    0    0    0    0    0    1
    0    0    0    -1    /exp;
estimate 'Small SD: Box vs Med' PartCond*Var*scenario_sd 0 -0.333333333    0    0    0    -0.333333333
    0    1    0    -0.333333333    0    0    /exp;
estimate 'Small SD: SD  vs Med' PartCond*Var*scenario_sd 0 -0.333333333    0    0    0    -0.333333333
    0    0    0    -0.333333333    0    1    /exp;
run; quit;
```

Analysis Of GEE Parameter Estimates									
Empirical Standard Error Estimates									
Parameter				**Estimate**	**Standard Error**	**95% Confidence Limits**		**Z**	**Pr > |Z|**
Intercept1				-1.8127	0.4817	-2.7568	-0.8685	-3.76	0.0002
Intercept2				-0.8965	0.4484	-1.7754	-0.0176	-2.00	0.0456
Intercept3				0.3040	0.4161	-0.5116	1.1195	0.73	0.4651
PartCon*Var*scenario	Dotplot	FALSE	large_5	1.3913	0.6024	0.2106	2.5720	2.31	0.0209
PartCon*Var*scenario	Dotplot	FALSE	small_2	1.9038	0.7162	0.5001	3.3075	2.66	0.0079
PartCon*Var*scenario	Dotplot	TRUE	large_5	0.7253	0.5352	-0.3237	1.7744	1.36	0.1754
PartCon*Var*scenario	Dotplot	TRUE	small_2	-0.2545	0.5139	-1.2618	0.7527	-0.50	0.6204
PartCon*Var*scenario	Boxplot	FALSE	large_5	0.6609	0.5862	-0.4880	1.8098	1.13	0.2595
PartCon*Var*scenario	Boxplot	FALSE	small_2	0.9019	0.5740	-0.2231	2.0270	1.57	0.1161
PartCon*Var*scenario	Boxplot	TRUE	large_5	1.0672	0.5717	-0.0533	2.1878	1.87	0.0619
PartCon*Var*scenario	Boxplot	TRUE	small_2	-0.4569	0.5779	-1.5896	0.6757	-0.79	0.4291
PartCon*Var*scenario	Standard_Deviation	FALSE	large_5	0.3098	0.3088	-0.2955	0.9151	1.00	0.3158

Analysis Of GEE Parameter Estimates

Empirical Standard Error Estimates

Parameter				Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
PartCon*Var*scenario	Standard_Deviation	FALSE	small_2	0.0139	0.4320	-0.8327	0.8605	0.03	0.9743
PartCon*Var*scenario	Standard_Deviation	TRUE	large_5	1.5642	0.4092	0.7622	2.3661	3.82	0.0001
PartCon*Var*scenario	Standard_Deviation	TRUE	small_2	0.0000	0.0000	0.0000	0.0000	.	.

Wald Statistics For Type 3 GEE Analysis

Source	DF	Chi-Square	Pr > ChiSq
PartCon*Var*scenario	11	101.22	<.0001

Contrast Estimate Results

Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Large vs. Small SD	0.6461	0.5877	0.7004	0.6018	0.1262	0.05	0.3544	0.8491	22.74	<.0001
Exp(Large vs. Small SD)				1.8253	0.2303	0.05	1.4254	2.3375		
Dotplot vs Boxplot	0.4826	0.2758	0.6955	-0.0697	0.4569	0.05	-0.9653	0.8258	0.02	0.8787
Exp(Dotplot vs Boxplot)				0.9326	0.4261	0.05	0.3809	2.2837		
Dotplot vs Mean&SD	0.3666	0.1928	0.5838	-0.5467	0.4515	0.05	-1.4317	0.3383	1.47	0.2260
Exp(Dotplot vs Mean&SD)				0.5789	0.2614	0.05	0.2389	1.4025		
Dotplot vs Median-only	0.3479	0.2052	0.5245	-0.6282	0.3705	0.05	-1.3544	0.0979	2.88	0.0900

Contrast Estimate Results										
		Mean					L'Beta			
Label	Mean Estimate	Confidence Limits		L'Beta Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
Exp(Dotplot vs Median-only)				0.5335	0.1977	0.05	0.2581	1.1029		
Boxplot vs Mean&SD	0.3830	0.1917	0.6189	-0.4770	0.4907	0.05	-1.4387	0.4848	0.94	0.3311
Exp(Boxplot vs Mean&SD)				0.6207	0.3046	0.05	0.2372	1.6239		
Boxplot vs Median-only	0.3639	0.2258	0.5287	-0.5585	0.3436	0.05	-1.2319	0.1150	2.64	0.1041
Exp(Boxplot vs Median-only)				0.5721	0.1966	0.05	0.2917	1.1218		
Mean&SD vs Median-only	0.4796	0.3226	0.6408	-0.0815	0.3370	0.05	-0.7420	0.5789	0.06	0.8088
Exp(Mean&SD vs Median-only)				0.9217	0.3106	0.05	0.4762	1.7841		
Dotplot: Large vs Small SD	0.7271	0.6604	0.7850	0.9799	0.1607	0.05	0.6649	1.2948	37.18	<.0001
Exp(Dotplot: Large vs Small SD)				2.6641	0.4281	0.05	1.9443	3.6503		
Boxplot: Large vs Small SD	0.8212	0.7248	0.8890	1.5242	0.2836	0.05	0.9683	2.0801	28.88	<.0001
Exp(Boxplot: Large vs Small SD)				4.5913	1.3023	0.05	2.6334	8.0051		
Mean&SD: Large vs Small SD	0.8270	0.6818	0.9142	1.5642	0.4092	0.05	0.7622	2.3661	14.61	0.0001
Exp(Mean&SD: Large vs Small SD)				4.7788	1.9553	0.05	2.1430	10.6563		
Median : Large vs Small SD	0.4619	0.3579	0.5694	-0.1526	0.2204	0.05	-0.5846	0.2795	0.48	0.4889

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Exp(Median : Large vs Small SD)				0.8585	0.1893	0.05	0.5573	1.3225		
Large SD: Dot vs Box	0.4153	0.2184	0.6436	-0.3419	0.4760	0.05	-1.2748	0.5910	0.52	0.4726
Exp(Large SD: Dot vs Box)				0.7104	0.3381	0.05	0.2795	1.8058		
Large SD: Dot vs SD	0.3018	0.1439	0.5263	-0.8389	0.4818	0.05	-1.7832	0.1055	3.03	0.0817
Exp(Large SD: Dot vs SD)				0.4322	0.2082	0.05	0.1681	1.1113		
Large SD: Dot vs Med	0.4845	0.3012	0.6721	-0.0620	0.3978	0.05	-0.8417	0.7177	0.02	0.8761
Exp(Large SD: Dot vs Med)				0.9399	0.3739	0.05	0.4310	2.0497		
Large SD: Box vs SD	0.3783	0.1835	0.6222	-0.4970	0.5081	0.05	-1.4928	0.4989	0.96	0.3280
Exp(Large SD: Box vs SD)				0.6084	0.3091	0.05	0.2247	1.6468		
Large SD: Box vs Med	0.5695	0.3870	0.7349	0.2799	0.3774	0.05	-0.4598	1.0196	0.55	0.4583
Exp(Large SD: Box vs Med)				1.3230	0.4993	0.05	0.6314	2.7721		
Large SD: SD vs Med	0.6850	0.5189	0.8143	0.7768	0.3577	0.05	0.0758	1.4779	4.72	0.0299
Exp(Large SD: SD vs Med)				2.1746	0.7778	0.05	1.0788	4.3836		
Small SD: Dot vs Box	0.5504	0.3205	0.7606	0.2024	0.4866	0.05	-0.7513	1.1561	0.17	0.6774

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Exp(Small SD: Dot vs Box)				1.2243	0.5958	0.05	0.4717	3.1777		
Small SD: Dot vs SD	0.4367	0.2207	0.6798	-0.2545	0.5139	0.05	-1.2618	0.7527	0.25	0.6204
Exp(Small SD: Dot vs SD)				0.7753	0.3984	0.05	0.2832	2.1227		
Small SD: Dot vs Med	0.2325	0.1219	0.3978	-1.1944	0.3979	0.05	-1.9743	-0.4145	9.01	0.0027
Exp(Small SD: Dot vs Med)				0.3029	0.1205	0.05	0.1389	0.6607		
Small SD: Box vs SD	0.3877	0.1694	0.6628	-0.4569	0.5779	0.05	-1.5896	0.6757	0.63	0.4291
Exp(Small SD: Box vs SD)				0.6332	0.3659	0.05	0.2040	1.9653		
Small SD: Box vs Med	0.1983	0.1002	0.3547	-1.3968	0.4073	0.05	-2.1952	-0.5985	11.76	0.0006
Exp(Small SD: Box vs Med)				0.2474	0.1008	0.05	0.1113	0.5496		
Small SD: SD vs Med	0.2809	0.1331	0.4986	-0.9399	0.4767	0.05	-1.8741	-0.0056	3.89	0.0486
Exp(Small SD: SD vs Med)				0.3907	0.1862	0.05	0.1535	0.9944		

Confidence in Predictions

```
proc mixed data = Exp2.data;
CLASS ParticipantID PartCond Var scenario_sd;
model EstimateConf = Var*PartCond*scenario_sd / residual ddfm=satterth solution;
random ParticipantID(PartCond);

estimate 'Large SD: Average Median-only' intercept 1 PartCond*Var*scenario_sd 0.333333333	0	0	0
	0.333333333	0	0	0	0.333333333	0	0	0 /cl;
estimate 'Large SD: Boxplot' intercept 1 PartCond*Var*scenario_sd 0 0	0	0	0	0	1
	0	0	0	0	0	/cl;
```

```
estimate 'Large SD: Dotplot' intercept 1 PartCond*Var*scenario_sd 0 0 1 0 0 0 0
    0 0 0 0 0 /cl;
estimate 'Large SD: Mean&SD' intercept 1 PartCond*Var*scenario_sd 0 0 0 0 0 0 0
    0 0 0 1 0 /cl;

estimate 'Small SD: Average Median-only' intercept 1 PartCond*Var*scenario_sd 0 0.333333333 0 0 0
    0.333333333 0 0 0 0.333333333 0 0 /cl;
estimate 'Small SD: Boxplot' intercept 1 PartCond*Var*scenario_sd 0 0 0 0 0 0 0
    1 0 0 0 0 /cl;
estimate 'Small SD: Dotplot' intercept 1 PartCond*Var*scenario_sd 0 0 0 1 0 0 0
    0 0 0 0 0 /cl;
estimate 'Small SD: Mean&SD' intercept 1 PartCond*Var*scenario_sd 0 0 0 0 0 0 0
    0 0 0 0 1 /cl;

estimate 'Large vs. Small SD' PartCond*Var*scenario_sd 0.166666667 -0.166666667 0.166666667 -0.166666667
0.166666667 -0.166666667 0.166666667 -0.166666667 0.166666667 -0.166666667 0.166666667 -0.166666667 /cl;

estimate 'Dotplot vs Boxplot' PartCond*Var*scenario_sd 0 0 0.5 0.5 0 0 -0.5 -0.5
    0 0 0 0 /cl;
estimate 'Dotplot vs Mean&SD' PartCond*Var*scenario_sd 0 0 0.5 0.5 0 0 0
    0 0 0 -0.5 -0.5 /cl;
estimate 'Dotplot vs Median-only' PartCond*Var*scenario_sd -0.166666667 -0.166666667 0.5 0.5 -
0.166666667 -0.166666667 0 0 -0.166666667 -0.166666667 0 0 /cl;
estimate 'Boxplot vs Mean&SD' PartCond*Var*scenario_sd 0 0 0 0 0 0 0.5
    0.5 0 0 -0.5 -0.5 /cl;
estimate 'Boxplot vs Median-only' PartCond*Var*scenario_sd -0.166666667 -0.166666667 0 0 -
0.166666667 -0.166666667 0.5 0.5 -0.166666667 -0.166666667 0 0 /cl;
estimate 'Mean&SD vs Median-only' PartCond*Var*scenario_sd -0.166666667 -0.166666667 0 0
    -0.166666667 -0.166666667 0 0 -0.166666667 -0.166666667 0.5 0.5 /cl;

estimate 'Dotplot: Large vs Small SD' PartCond*Var*scenario_sd 0 0 1 -1 0 0 0
    0 0 0 0 0 / cl;
estimate 'Boxplot: Large vs Small SD' PartCond*Var*scenario_sd 0 0 0 0 0 0 1
    -1 0 0 0 0 / cl;
estimate 'Mean&SD: Large vs Small SD' PartCond*Var*scenario_sd 0 0 0 0 0 0
    0 0 0 0 1 -1 /cl;
estiamte 'Median : Large vs Small SD' PartCond*Var*scenario_sd 0.333333333 -0.333333333 0 0
    0.333333333 -0.333333333 0 0 0.333333333 -0.333333333 0 0 / cl;

estimate 'Large SD: Dot vs Box' PartCond*Var*scenario_sd 0 0 1 0 0 0 -1
    0 0 0 0 0 / cl;
estimate 'Large SD: Dot vs SD ' PartCond*Var*scenario_sd 0 0 1 0 0 0 0
    0 0 0 -1 0 / cl;
estimate 'Large SD: Dot vs Med' PartCond*Var*scenario_sd -0.333333333 0 1 0 -0.333333333
    0 0 0 -0.333333333 0 0 0 /cl;
estimate 'Large SD: Box vs SD ' PartCond*Var*scenario_sd 0 0 0 0 0 0 1
    0 0 0 -1 0 /cl;
estimate 'Large SD: Box vs Med' PartCond*Var*scenario_sd -0.333333333 0 0 0 -0.333333333
    0 1 0 -0.333333333 0 0 0 /cl;
estimate 'Large SD: SD  vs Med' PartCond*Var*scenario_sd -0.333333333 0 0 0 -0.333333333
    0 0 0 -0.333333333 0 1 0 /cl;

estimate 'Small SD: Dot vs Box' PartCond*Var*scenario_sd 0 0 0 1 0 0 0 -1
    0 0 0 0 / cl;
estimate 'Small SD: Dot vs SD ' PartCond*Var*scenario_sd 0 0 0 1 0 0 0 0
    0 0 0 -1 / cl;
estimate 'Small SD: Dot vs Med' PartCond*Var*scenario_sd 0 -0.333333333 0 1 0 -0.333333333
    0 0 0 -0.333333333 0 0 /cl;
estimate 'Small SD: Box vs SD ' PartCond*Var*scenario_sd 0 0 0 0 0 0 0 1
    0 0 0 -1 /cl;
estimate 'Small SD: Box vs Med' PartCond*Var*scenario_sd 0 -0.333333333 0 0 0 -0.333333333
    0 1 0 -0.333333333 0 0 /cl;
```

```
estimate 'Small SD: SD  vs Med' PartCond*Var*scenario_sd 0 -0.333333333    0           0           0           -0.333333333
            0           0           0           -0.333333333       0           1           /cl;
run; quit;
```

Solution for Fixed Effects

Effect	PartCond	Var	scenario_sd	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept				65.7833	4.2933	74.2	15.32	<.0001
PartCon*Var*scenario	Dotplot	FALSE	2	-3.8333	6.0717	74.2	-0.63	0.5298
PartCon*Var*scenario	Dotplot	FALSE	5	-4.2667	6.0717	74.2	-0.70	0.4844
PartCon*Var*scenario	Dotplot	TRUE	2	1.4000	6.0717	74.2	0.23	0.8183
PartCon*Var*scenario	Dotplot	TRUE	5	-15.9000	6.0717	74.2	-2.62	0.0107
PartCon*Var*scenario	Boxplot	FALSE	2	-10.5833	6.0717	74.2	-1.74	0.0855
PartCon*Var*scenario	Boxplot	FALSE	5	-9.6000	6.0717	74.2	-1.58	0.1181
PartCon*Var*scenario	Boxplot	TRUE	2	-6.9667	6.0717	74.2	-1.15	0.2549
PartCon*Var*scenario	Boxplot	TRUE	5	-14.3000	6.0717	74.2	-2.36	0.0212
PartCon*Var*scenario	Standard_Deviation	FALSE	2	-1.7000	2.4712	651	-0.69	0.4917
PartCon*Var*scenario	Standard_Deviation	FALSE	5	-4.7500	2.4712	651	-1.92	0.0550
PartCon*Var*scenario	Standard_Deviation	TRUE	2	4.3167	2.4712	651	1.75	0.0811
PartCon*Var*scenario	Standard_Deviation	TRUE	5	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
PartCon*Var*scenario	11	238	7.14	<.0001

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Large SD: Average Median-only	60.4111	2.4788	74.2	24.37	<.0001	0.05	55.4723	65.3499

Estimates								
Label	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Large SD: Boxplot	58.8167	4.2933	74.2	13.70	<.0001	0.05	50.2624	67.3709
Large SD: Dotplot	67.1833	4.2933	74.2	15.65	<.0001	0.05	58.6291	75.7376
Large SD: Mean&SD	70.1000	4.2933	74.2	16.33	<.0001	0.05	61.5457	78.6543
Small SD: Average Median-only	59.5778	2.4788	74.2	24.04	<.0001	0.05	54.6390	64.5166
Small SD: Boxplot	51.4833	4.2933	74.2	11.99	<.0001	0.05	42.9291	60.0376
Small SD: Dotplot	49.8833	4.2933	74.2	11.62	<.0001	0.05	41.3291	58.4376
Small SD: Mean&SD	65.7833	4.2933	74.2	15.32	<.0001	0.05	57.2291	74.3376
Large vs. Small SD	5.2417	1.0089	651	5.20	<.0001	0.05	3.2607	7.2227
Dotplot vs Boxplot	3.3833	5.8148	62.5	0.58	0.5628	0.05	-8.2384	15.0051
Dotplot vs Mean&SD	-9.4083	5.8148	62.5	-1.62	0.1107	0.05	-21.0301	2.2134
Dotplot vs Median-only	-1.4611	3.5055	74.2	-0.42	0.6780	0.05	-8.4456	5.5234
Boxplot vs Mean&SD	-12.7917	5.8148	62.5	-2.20	0.0315	0.05	-24.4134	-1.1699
Boxplot vs Median-only	-4.8444	3.5055	74.2	-1.38	0.1711	0.05	-11.8290	2.1401
Mean&SD vs Median-only	7.9472	3.5055	74.2	2.27	0.0263	0.05	0.9627	14.9318
Dotplot: Large vs Small SD	17.3000	2.4712	651	7.00	<.0001	0.05	12.4475	22.1525
Boxplot: Large vs Small SD	7.3333	2.4712	651	2.97	0.0031	0.05	2.4809	12.1858
Mean&SD: Large vs Small SD	4.3167	2.4712	651	1.75	0.0811	0.05	-0.5358	9.1691
Median : Large vs Small SD	0.8333	1.4267	651	0.58	0.5594	0.05	-1.9682	3.6349
Large SD: Dot vs Box	8.3667	6.0717	74.2	1.38	0.1724	0.05	-3.7309	20.4642
Large SD: Dot vs SD	-2.9167	6.0717	74.2	-0.48	0.6324	0.05	-15.0142	9.1809
Large SD: Dot vs Med	6.7722	3.7847	100	1.79	0.0766	0.05	-0.7365	14.2810
Large SD: Box vs SD	-11.2833	6.0717	74.2	-1.86	0.0671	0.05	-23.3809	0.8142

Estimates								
Label	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Large SD: Box vs Med	-1.5944	3.7847	100	-0.42	0.6745	0.05	-9.1032	5.9143
Large SD: SD vs Med	9.6889	3.7847	100	2.56	0.0120	0.05	2.1802	17.1976
Small SD: Dot vs Box	-1.6000	6.0717	74.2	-0.26	0.7929	0.05	-13.6976	10.4976
Small SD: Dot vs SD	-15.9000	6.0717	74.2	-2.62	0.0107	0.05	-27.9976	-3.8024
Small SD: Dot vs Med	-9.6944	3.7847	100	-2.56	0.0119	0.05	-17.2032	-2.1857
Small SD: Box vs SD	-14.3000	6.0717	74.2	-2.36	0.0212	0.05	-26.3976	-2.2024
Small SD: Box vs Med	-8.0944	3.7847	100	-2.14	0.0349	0.05	-15.6032	-0.5857
Small SD: SD vs Med	6.2056	3.7847	100	1.64	0.1042	0.05	-1.3032	13.7143

<u>Prediction Probability</u>

On Central Tendency

```
proc mixed data = Exp2.data_nondev;
CLASS ParticipantID PartCond Var scenario_sd;
model Prob = Var*PartCond / residual ddfm=satterth solution outp=res ;
repeated / subject = ParticipantID(PartCond) type = cs;
estimate 'Dotplot' intercept 1 PartCond*Var 0 1 0 0 0 0 / cl;
estimate 'Boxplot' intercept 1 PartCond*Var 0 0 0 1 0 0 / cl;
estimate 'Mean&SD' intercept 1 PartCond*Var 0 0 0 0 0 1 / cl;
estimate 'Med-only' intercept 1 PartCond*Var 0.333333333 0    0.333333333    0    0.333333333    0 / cl;
estimate 'Dot vs Box' PartCond*Var 0   1    0    -1    0    0 / cl;
estimate 'Dot vs SD ' PartCond*Var 0   1    0    0    0    -1 / cl;
estimate 'Dot vs Med' PartCond*Var -0.333333333    1    -0.333333333    0    -0.333333333    0 /cl;
estimate 'Box vs SD ' PartCond*Var 0   0    0    1    0    -1 /cl;
estimate 'Box vs Med' PartCond*Var -0.333333333    0    -0.333333333    1    -0.333333333    0 /cl;
estimate 'Dot vs Box' PartCond*Var -0.333333333    0    -0.333333333    0    -0.333333333    1 /cl;
run; quit;
```

Solution for Fixed Effects							
Effect	**PartCond**	**Var**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept			57.7725	3.8461	39.2	15.02	<.0001
PartCond*Var	Dotplot	FALSE	-3.2173	5.0009	39.1	-0.64	0.5237

Solution for Fixed Effects							
Effect	**PartCond**	**Var**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
PartCond*Var	Dotplot	TRUE	-5.5931	5.3160	49.2	-1.05	0.2979
PartCond*Var	Boxplot	FALSE	-4.9098	5.4995	37.5	-0.89	0.3777
PartCond*Var	Boxplot	TRUE	-6.3067	5.6032	40.3	-1.13	0.2670
PartCond*Var	Standard_Deviation	FALSE	-4.4809	2.3389	232	-1.92	0.0566
PartCond*Var	Standard_Deviation	TRUE	0	.	.	.	.

Type 3 Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
PartCond*Var	5	72	0.99	0.4289

Estimates								
Label	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Dotplot	52.1793	3.6698	64.6	14.22	<.0001	0.05	44.8494	59.5093
Boxplot	51.4658	4.0748	41.3	12.63	<.0001	0.05	43.2385	59.6930
Mean&SD	57.7725	3.8461	39.2	15.02	<.0001	0.05	49.9944	65.5505
Med-only	53.5698	2.0986	36.4	25.53	<.0001	0.05	49.3153	57.8243
Dot vs Box	0.7136	5.4837	50	0.13	0.8970	0.05	-10.3008	11.7279
Dot vs SD	-5.5931	5.3160	49.2	-1.05	0.2979	0.05	-16.2751	5.0888
Dot vs Med	-1.3905	3.5672	87.7	-0.39	0.6976	0.05	-8.4799	5.6989
Box vs SD	-6.3067	5.6032	40.3	-1.13	0.2670	0.05	-17.6285	5.0151
Box vs Med	-2.1041	3.5403	60.8	-0.59	0.5545	0.05	-9.1839	4.9758
Dot vs Box	4.2027	3.3813	55.5	1.24	0.2191	0.05	-2.5724	10.9777

Not On Central Tendency

```
proc mixed data = Exp2.dev;
CLASS ParticipantID PartCond Var scenario_sd;
model Prob = Var*PartCond/ residual ddfm=satterth solution outp=res ;
random intercept/ subject = ParticipantID type = cs;
estimate 'Dotplot' intercept 1 PartCond*Var 0 1 0 0 0 0 / cl;
estimate 'Boxplot' intercept 1 PartCond*Var 0 0 0 1 0 0 / cl;
estimate 'Mean&SD' intercept 1 PartCond*Var 0 0 0 0 0 1 / cl;
estimate 'Med-only' intercept 1 PartCond*Var 0.333333333 0    0.333333333    0    0.333333333    0 / cl;
estimate 'Dot vs Box' PartCond*Var 0   1    0    -1    0    0 / cl;
estimate 'Dot vs SD ' PartCond*Var 0   1    0    0    0    -1 / cl;
estimate 'Dot vs Med' PartCond*Var -0.333333333    1    -0.333333333    0    -0.333333333    0 /cl;
estimate 'Box vs SD ' PartCond*Var 0   0    0    1    0    -1 /cl;
estimate 'Box vs Med' PartCond*Var -0.333333333    0    -0.333333333    1    -0.333333333    0 /cl;
estimate 'Dot vs Box' PartCond*Var -0.333333333    0    -0.333333333    0    -0.333333333    1 /cl;
run; quit;
```

Solution for Fixed Effects

Effect	PartCond	Var	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			65.5288	3.2389	52.7	20.23	<.0001
PartCond*Var	Dotplot	FALSE	-2.1603	4.5328	63	-0.48	0.6353
PartCond*Var	Dotplot	TRUE	-5.5578	4.3061	52.7	-1.29	0.2024
PartCond*Var	Boxplot	FALSE	-5.3874	4.4666	56.4	-1.21	0.2328
PartCond*Var	Boxplot	TRUE	-5.8410	4.3938	53.5	-1.33	0.1894
PartCond*Var	Standard_Deviation	FALSE	1.4809	2.2174	410	0.67	0.5046
PartCond*Var	Standard_Deviation	TRUE	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
PartCond*Var	5	103	1.00	0.4219

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Dotplot	59.9711	2.8376	52.7	21.13	<.0001	0.05	54.2789	65.6633
Boxplot	59.6879	2.9690	54.5	20.10	<.0001	0.05	53.7365	65.6393

Estimates								
Label	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Mean&SD	65.5288	3.2389	52.7	20.23	<.0001	0.05	59.0315	72.0261
Med-only	63.5066	1.8483	64.4	34.36	<.0001	0.05	59.8146	67.1985
Dot vs Box	0.2832	4.1070	53.6	0.07	0.9453	0.05	-7.9521	8.5185
Dot vs SD	-5.5578	4.3061	52.7	-1.29	0.2024	0.05	-14.1958	3.0803
Dot vs Med	-3.5355	2.7063	83.8	-1.31	0.1950	0.05	-8.9175	1.8465
Box vs SD	-5.8410	4.3938	53.5	-1.33	0.1894	0.05	-14.6520	2.9700
Box vs Med	-3.8187	2.7611	84.9	-1.38	0.1703	0.05	-9.3087	1.6713
Dot vs Box	2.0223	2.8820	84.2	0.70	0.4848	0.05	-3.7086	7.7531

Appendix E – Pilot Experiment on Visualization Effects on Estimation Behavior

This Appendix was first presented at the 12th International Naturalistic Decision Making Conference (Giang & Donmez, 2015).

Interpreting Visualizations of Historical Variability for Estimating Future Events

Wayne Chi Wei GIANG and Birsen DONMEZ
Human Factors and Applied Statistics Laboratory, Mechanical and Industrial Engineering, University of Toronto

ABSTRACT
Estimations of variables like travel time are often important for scheduling and logistics decisions. When these estimations are made under high time pressure, visualizations of historical data can be used to help produce more accurate estimates and decisions. In this preliminary study, we examine four visualizations that represent increasing amounts of information about the dispersion and shape of the historical data and examine how these visualizations are used to produce estimates of task completion times. In particular, we are interested in whether the level of variability information provided causes the decision makers to systematically adjust their estimate away from the central tendency of the historical data. We found that participants were more confident and tended to deviate from the central tendency more when they had information about the range and shape of historical data compared to when they only had a point estimate or a point estimate and standard deviation.

KEYWORDS
Planning and prediction; transportation; estimation; visualizations.

INTRODUCTION

Humans are often required to make decisions in situations that are characterized by large degrees of complexity and uncertainty that cannot be deterministically modelled. These situations often arise due to incomplete or uncertain information about the current or future state of the world, and thus decision makers must estimate and predict the variables that are critical for their decisions. Furthermore, the presence of time pressure increases the difficulty of these decisions and may lead to the use of heuristics and biases, which may not always be appropriate (Payne et al., 1993; Tversky & Kahneman, 1974). Decision support systems providing historical data trends is one method to support evidence-based decisions that can help mitigate some of these difficulties.

An example scenario where decision makers are required to make short-term, time-critical, evidence-based decisions is medical dispatch. For example, medical dispatchers at Ornge, the medical transport system in Ontario, Canada, make their dispatch decisions by estimating time to definitive care, i.e., how long it takes to transfer patients between hospitals. However, in our previous work (Giang, Donmez, Fatahi, Ahghari, & MacDonald, 2014), we had identified that Ornge dispatcher estimates of time to definitive care tended to be shorter than actual times. Transfers are impacted by a variety of factors, such as traffic and patient condition. These factors are often hard for dispatchers to account for, and can cause transfers to deviate from normally expected times.

In response to these findings, a decision support tool was developed to provide the dispatchers with estimates based on historical transfer information that had less error than the dispatchers' own estimates. The tool provided point-estimates of transfer times calculated based on descriptive statistics and linear models. However, the point-estimates only communicated information on central tendency and fail to provide insight about the historical variability and uncertainty associated with these estimates. Ornge's dispatchers are expert decision makers who often have additional contextual information (e.g., knowledge about the crews involved or the weather) that they use to modify their own transfer time estimates. Visualizations of the variability of historical data may allow disaptchers to use this contexutal information in a way that is tied to historical data.

However, there has been little work done on how visualizations of historical data are interpreted for decision making. Uncertainty visualizations for dynamic decision making scenarios have typically dealt with providing classification information about objects instead of display information about the variability of continuous variables. For example, Neyedli, Hollands, and Jamieson (Neyedli et al., 2011) developed and tested visualizations that showed the reliability of a system which detects friendly or enemy targets. Bisantz et al. (2011) developed visualizations of the uncertainty associated with object classifications in a missile detection game. In both these examples, the data that is being visualized is a classification of an object, the uncertainty or reliability information is a measurement of the likelihood of belonging to a category, and decision makers must use this information to make a judgment of the true identity of the object. However in applications such as medical dispatch, a critical part of logistics and coordination is the estimation of a specific time as opposed to a judgment of which category (e.g., late or not late) the time estimation belongs. Tasks such as scheduling ambulance arrival times, booking helipads, and letting staff at the receiving facility know when they should expect to receive a patient all benefit from having more accurate time estimates.

Presenting uncertainty information about continuous variables has shown mixed results in terms of performance and usage that appear to be highly tied to the method of presentation. Nadav-Greenberg and Joslyn (2009) examined verbal, numeric, and graphical representations of uncertainty information about nighttime temperature lows in a road-salting decision task. Participants were asked to predict the expected nighttime low, while making a decision about whether to salt the road if the temperature was expected to drop below freezing. They found that the uncertainty information helped participants make better decisions about road salting, and that the estimates of the nighttime lows were impacted by the type of information given (e.g., full range, probability of freezing). However, they only explored one graphical representation of uncertainty (for full range only) which they found to be not as effective in improving salting decisions in comparison to numerical representations. Similarly, Scown, Bartlett, and McCarley (2014) found that non-expert decision makers often did not use error bars when making two-point comparisons about product review scores. The benefits of visual representations of uncertainty information are often harder to study because individuals tend to construct different internal models of underlying probability distributions that are influenced by the graphical elements of the visualization (Tak, Toet, & van Erp, 2014). Thus, there may be factors that influence how visualizations of variability information are interpreted by decision makers, and these effects might be tied to the amount of variability information provided.

The goal of this preliminary study is to examine how the amount of variability information influences the way decision makers interpret visualizations of historical data. We examined four visualizations that represent increasing amounts of information about the dispersion and shape of the historical data: central tendency only, mean and standard deviation, boxplot, and violin plot. In particular, we are interested in examining whether the level of variability information provided will cause the decision makers to systematically adjust their estimate away from the central tendency (i.e., median, mode, or mean) of the historical data.

METHODS

Participants and Apparatus

We recruited 22 participants from the local community and the undergraduate and graduate population at the University of Toronto. Participants were selected using a screening questionnaire for completion of at least one probability or statistics course during their post-secondary education. Furthermore, all participants had normal or corrected-to-normal vision and normal colour perception.

Of the 22 participants, 13 were male and 9 were female. Participant ages ranged between 19 and 30 with a mean (M) of 24.5 years and a standard deviation (SD) of 3.0 years. Participants also reported taking an average of 1.8 probability or statistics courses during their post-secondary education (SD = 0.8), with 10 participants having taken these courses in graduate school and 12 participants at the undergraduate level.

The experiment was conducted in a quiet office environment. Participants were seated in front of a 24-inch monitor that displayed the experimental tasks. Participants responded to the tasks using a keyboard and mouse. The experimental software was created using the open-source PsychoPy framework (Pierce, 2007).

Experimental Scenario and Task

An experimental scenario was created where the participants would not have any contextual information to draw from other than the information presented in the visualizations. Participants took on the role of a mission commander responsible to overseeing a number of scientific space rovers exploring a planet. The role of the mission commander was to monitor the amount of time required for a rover to complete a scientific task (e.g., collecting or analysing samples) in order to determine whether the rover would be able to stay on schedule with their upcoming tasks.

Participants were told that each rover had their own set of historical data that represented the task completion times for that rover in the past, so that every rover should be treated independently.

Eight datasets were generated to serve as the historical data for the rovers. Each of these datasets was formed by sampling 50 data points from normal distributions with 4 different means and 2 levels of standard deviation for each mean. The four means used were 33, 56, 70, and 79, and the two levels of standard deviation were 10% of the mean and 30% of the mean. In addition, a "true" task time was also sampled from the distribution which represented the correct task completion time. The 8 datasets were presented using each of the 4 visualizations and replicated 4 times for a total of 128 trials per participant. Each of these trials represented a new rover that the participants had to monitor.

Participants were responsible for two tasks, similar to those used by Nadav-Greenberg and Joslyn (2009). In the first task, i.e., the estimation task, participants were required to estimate how long they thought it would take for the rover to complete its current scientific task. The participants selected a value using a slider scale that was superimposed on the uncertainty visualization, as shown in Figure 1. Estimates were restricted to integer values. In the second task, i.e., the judgment task, participants were asked to make a judgment about whether they felt that the rover was going to be able to complete its task by a specific cut-off time, also shown in Figure 1. Participants were also asked to rate their confidence in these two tasks on a scale between 1 and 100. During the experiment, participants completed trials containing both the estimation task and the judgment task, with the order of task presentation counterbalanced.

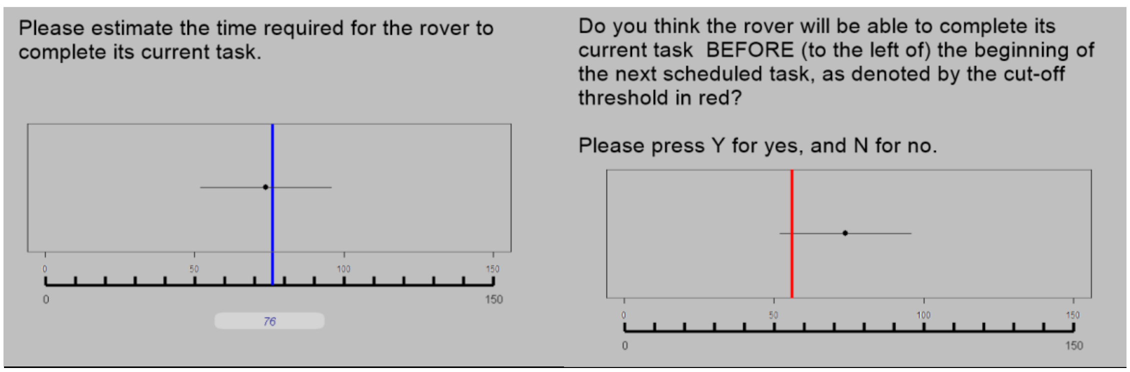


Figure 1: The estimation task with a slider bar (left) and the judgment task with the cut-off time (right).

Experimental Design

The primary independent variable of interest was the type of visualization of the historical data, a within subject variable. Four visualization conditions (Figure 2) were designed for this experiment with increasing amounts of information about the dispersion and shape of the distribution of historical observations. In a baseline, central tendency only condition, only the median of the historical sample was displayed. The mean & standard deviation visualization condition showed both a measure of central tendency and a measure of dispersion but did not provide information about the shape of the historical sample. The boxplot visualization provided a measure of central tendency (median) as well as two measures of dispersion (interquartile range and range). The boxplot visualization also provided information about the skewness of the historical sample; the kurtosis of the historical sample could be inferred from the relative lengths of the box and whiskers. Finally, the violin plot visualization provided an indicator for the median, the interquartile range and range, as well as an estimate of the distribution of the historical sample (i.e., a kernel density estimate) which provides information of both skewness and kurtosis. All visualizations were generated using R, with the violin plots created using the 'vioplot' package. The x-axis in these visualizations represented minutes (task completion time), while the y-axis in the violin plot represented an estimate of the probability density of the historical sample.

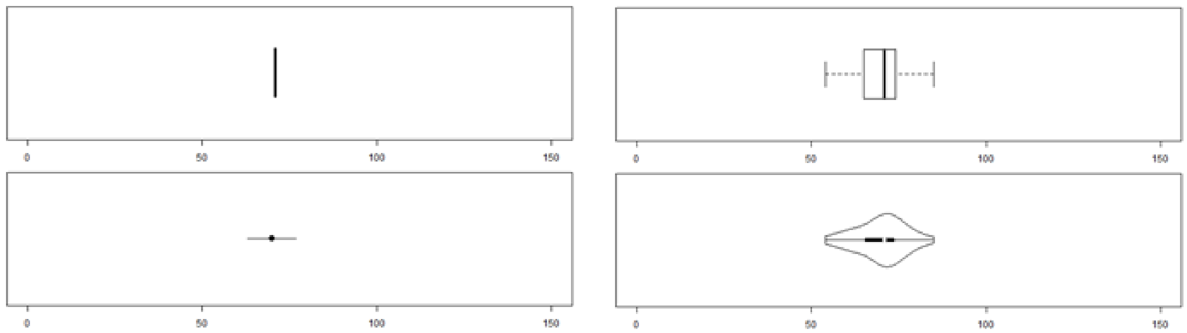


Figure 2: The visualizations of historical data: a) median only, b) mean & standard deviation, c) boxplot, and d) violin plot.

Experimental Procedure

Before beginning the experiment, participants were given a review of the concepts of central tendency and dispersion, and an introduction to the visualizations used in the experiment. Participants were then provided with a short practice session of 4 trials for each of the visualization conditions. They were also told that there would be a $5 performance bonus during the experimental trials, although everyone was given this performance bonus.

In the first half of the experiment, participants were not given any feedback on their task performance. In the second half, participants were provided with feedback about the "true" task duration for that rover, and feedback about the correctness of their judgments after each trial. The experiment took approximately 90 minutes to complete, and participants were paid $20, which included the performance bonus.

Data Processing

The analysis for this paper focuses on the no feedback trials. There were three dependent variables of interest with respect to the estimation of task times: 1) the distance between the participant's estimate and the central tendency expressed in number of standard deviations of the historical data, 2) the participant's confidence in the estimate rated between 1 and 100, and 3) the number of times the participant's estimate was not a central tendency point. The first dependent variable was calculated by using the closest measure of central tendency (i.e., mode, median, or mean) to the participant's estimate since each visualization type had a different prominently displayed central tendency measure, and for the violin plot, there were multiple features that the participant could use as a central tendency point. Since participants were restricted to responding in integer values, a correction was applied for the third dependent variable: a response was counted to be on a central tendency measure as long as it fell within 0.5 units below or above it. Seven outlying trials, which had participant estimates greater than 2 standard deviations from the closest central tendency were removed. For all three dependent variables, values were averaged for each participant up to the visualization level.

RESULTS

A linear mixed model was fitted to the distance to central tendency data (dependent variable 1) as a function of visualization type (median, mean & standard deviation, boxplot, and violin plot). Participant was used as a random factor and a square-root transformation was employed to ensure normality of residuals. Visualization type was found to be a significant factor, $F(3, 63) = 3.33$, $p = .03$. Post-hoc analyses using Tukey contrasts found that the violin plot ($\Delta = 0.074$, $p = .02$) had significantly larger distances to central tendency compared to the median only visualization, while the boxplot ($\Delta = 0.063$, $p = .06$) had only marginally significantly larger distances to central tendency than the median only visualization.

A second linear mixed model was fitted to the confidence data as a function of visualization type, with participant as a random factor. Visualization type was found to be significant, $F(3, 63) = 17.6$, $p < .0001$. Post-hoc comparisons using Tukey contrasts revealed that the median only visualization resulted in significantly lower confidence ratings than the mean & standard deviation ($\Delta = 5.7$, $p < .001$), the boxplot ($\Delta = 7.1$, $p < .001$), and the violin plot ($\Delta = 9.3$, $p < .001$) visualizations. In addition, the mean & standard deviation visualization also resulted in lower confidence ratings than the violin plot, $\Delta = 3.5$, $p = .04$.

Finally, a Poisson regression model was fitted to the third dependent variable (number of times the participant's estimate was not a central tendency point) with visualization type as a predictor variable. The number of trials completed by each participant per visualization (typically 16) was used as an offset variable, and participant was treated as a random factor. Again, visualization type was found to be a significant factor, $\chi^2(3) = 33.3$. $p < .001$. The median data values for this dependent variable for the four visualizations (i.e., median only, mean & standard deviation, boxplot, and violin plot) were 4.5, 5.5, 10, and 11, respectively. Post-hoc analysis using Tukey contrasts found that participants had a greater number of deviations from the central tendency with the boxplot ($p < .001$) and the violin plot ($p < .001$) visualizations when compared to the median only visualization. Similarly, the boxplot ($p = .003$) and the violin plot ($p < .001$) visualizations also resulted in a greater number of deviations than the the mean & standard deviation visualization.

DISCUSSION AND CONCLUSION

These results suggest that the type of visualization used has an impact on how individuals make estimations through historical data. Participants did deviate away from the central tendency, and deviations were both more likely to occur and also to have larger magnitudes for the visualizations that provided more variability information (i.e., boxplot and violin plot). Participants' ratings of confidence were also higher for the boxplot and the violin plot visualizations. The central tendency points are the statistically optimal responses (since the scenarios were sampled from normal distributions), yet it appeared that the shape and range information led our participants to feel safe about estimating away from central tendency. It is also possible that participants had a harder time picking out a single measure of central tendency from the graphs containing more information, leading to the larger deviations from the central tendency measures that we calculated. Overall, this preliminary study suggests that participants do try to use variability information in adjusting their estimates of future events using historical data, and that further study of how contextual information might also influence their estimates is required.

ACKNOWLEDGMENTS

We thank the Natural Sciences and Engineering Research Council and Ornge for their funding and support.

REFERENCES

Bisantz, A. M., Cao, D., Jenkins, M., Pennathur, P. R., Farry, M., Roth, E., Potter, S. S., Pfautz, J. (2011). Comparing uncertainty visualizations for a dynamic decision-making task. *Journal of Cognitive Engineering and Decision Making*, 5(3), 277–293.

Giang, W. C. W., Donmez, B., Fatahi, A., Ahghari, M., & MacDonald, R. D. (2014). Supporting Air Versus Ground Vehicle Decisions for Interfacility Medical Transport Using Historical Data. *IEEE Transactions on Human-Machine Systems*, 44(1), 55–65.

Nadav-Greenberg, L., & Joslyn, S. L. (2009). Uncertainty Forecasts Improve Decision Making Among Nonexperts. *Journal of Cognitive Engineering and Decision Making*, 3(3), 209–227.

Neyedli, H. F., Hollands, J. G., & Jamieson, G. A. (2011). Beyond Identity: Incorporating System Reliability Information Into an Automated Combat Identification System. Human Factors: *The Journal of the Human Factors and Ergonomics Society*, 53(4), 338–355.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. New York, NY: Cambridge University Press.

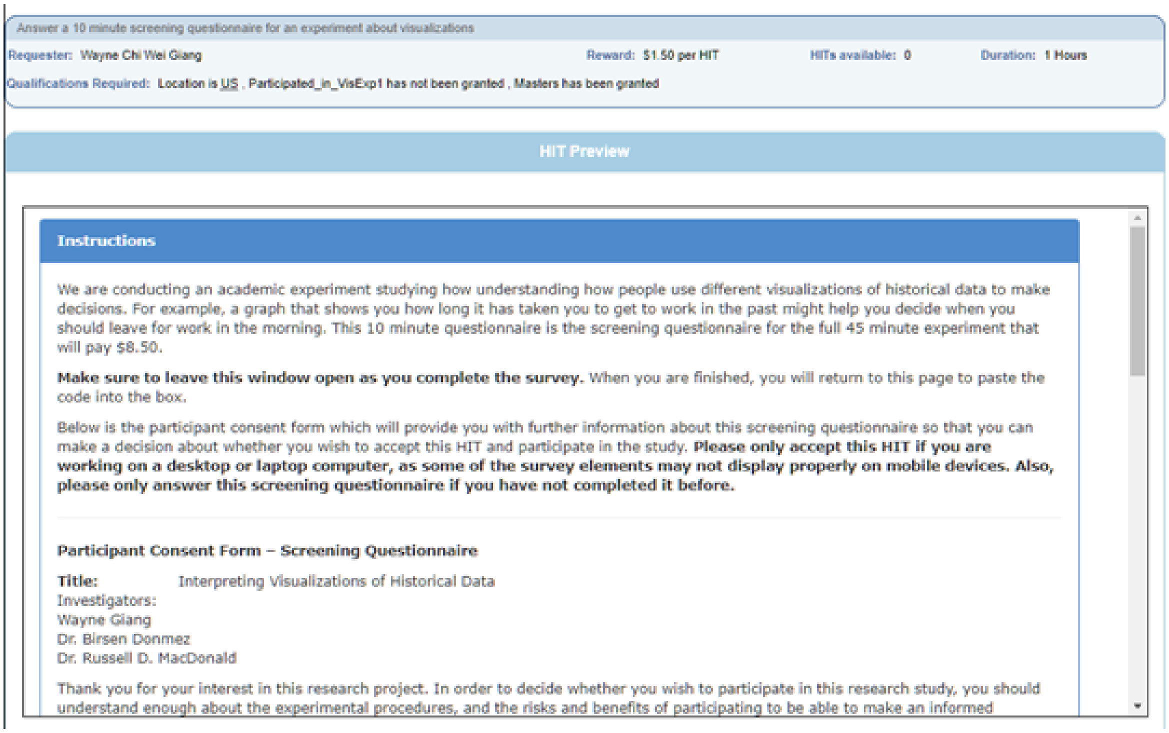
Scown, H., Bartlett, M., & McCarley, J. S. (2014). Statistically Lay Decision Makers Ignore Error Bars in Two-Point Comparisons. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 1746–1750.

Tak, S., Toet, A., & Erp, J. Van. (2014). The Perception of Visual Uncertainty Representation by Non-Experts. *IEEE Transactions on Visualization and Computer Graphics*, 20(6), 935–943.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–31.

Appendix F – Experimental Materials for Chapter 7

Amazon Mechanical Turk HIT for Screening Questionnaire



Answer a 10 minute screening questionnaire for an experiment about visualizations

Requester: Wayne Chi Wei Giang | Reward: $1.50 per HIT | HITs available: 0 | Duration: 1 Hours

Qualifications Required: Location is US , Participated_in_VisExp1 has not been granted , Masters has been granted

HIT Preview

Instructions

We are conducting an academic experiment studying how understanding how people use different visualizations of historical data to make decisions. For example, a graph that shows you how long it has taken you to get to work in the past might help you decide when you should leave for work in the morning. This 10 minute questionnaire is the screening questionnaire for the full 45 minute experiment that will pay $8.50.

Make sure to leave this window open as you complete the survey. When you are finished, you will return to this page to paste the code into the box.

Below is the participant consent form which will provide you with further information about this screening questionnaire so that you can make a decision about whether you wish to accept this HIT and participate in the study. **Please only accept this HIT if you are working on a desktop or laptop computer, as some of the survey elements may not display properly on mobile devices. Also, please only answer this screening questionnaire if you have not completed it before.**

Participant Consent Form – Screening Questionnaire

Title: Interpreting Visualizations of Historical Data
Investigators:
Wayne Giang
Dr. Birsen Donmez
Dr. Russell D. MacDonald

Thank you for your interest in this research project. In order to decide whether you wish to participate in this research study, you should understand enough about the experimental procedures, and the risks and benefits of participating to be able to make an informed

Consent Form for Screening Questionnaire

Visualization Experiment Screening Questionnaire

Participant Consent Form

Mechanical & Industrial Engineering
UNIVERSITY OF TORONTO

Participant Consent Form – Screening Questionnaire

Title: Interpreting Visualizations of Historical Data
Investigators:
Wayne Giang
Dr. Birsen Donmez
Dr. Russell D. MacDonald

Thank you for your interest in this research project. In order to decide whether you wish to participate in this research study, you should understand enough about the experimental procedures, and the risks and benefits of participating to be able to make an informed decision. This is known as the informed consent process.

If at any point you feel as though any of the following details are unclear, or if you have any other questions, comments, or concerns, please feel free to contact me using the contact information provided above. **If you decide that you would like to participate, please check off all checkboxes at the bottom of this page. If you do not wish to participate, there is no need to return the form and you may close this window and return the HIT.**

Purpose
We are interested in understanding how people use different visualizations of historical data to make decisions. For example, a graph that shows you how long it has taken you to get to work in the past might help you decide when you should leave for work in the morning.

This current survey is the screening questionnaire for the experiment, and if you qualify you may be invited back to participate in the experiment.

As a participant you will be asked to:

- Complete two questionnaires that relate to personality factors and demographics.
- Read about concepts that are important in understanding visualizations of historical data, and respond to a short 8-question test on these concepts.

Procedure
There are two parts to this survey. In the first part you will be asked to fill out two questionnaires that relate to demographics data and personality factors. The first questionnaire asks for demographics data that may be related to understanding visualizations of historical data such as age, profession, and math education. The second questionnaire examines how likely you are to use numbers in your everyday life. In the second part, you will be presented with information about how historical data is summarized and described. You will be asked to answer a short 8-question test on these concepts.

Risks
There are no major risks involved with this experiment; the tasks are not physiologically demanding, or psychologically stressing.

Benefits
There are several benefits to conducting this study. The most important benefit is your contribution to research in information visualization and decision support systems, which will guide the development of future decision support systems for expert decision makers.

Compensation
This survey should take approximately 10 minutes, and you will be compensated $1.50 upon completion of all the questions. **Based on the results of this survey, you may be invited to participate in a 45 minute experiment, which will be compensated $8.50.**

Confidentiality
All information obtained during the study will be held in strict confidence. You will be identified with a study number only, and this study number will only be identifiable by the researchers. No names or identifying information will be used in any publication or presentation. No information identifying you will be transferred outside the investigators in this study.

Participation
Your participation in this study is voluntary. You can choose to not participate or withdraw at any time by closing this window and returning the HIT before completing all the questions. If you choose to withdraw, you will not be compensated for the study and your data will be deleted.

Questions
If you have any general questions about this study, please call 416.978.0881 or email wayne.giang@mail.utoronto.ca. You can also contact the Office of Research Ethics at ethics.review@utoronto.ca or 416-946-3273, if you have questions about your rights as participants.

1. To be completed by participants, please check all three if you agree to participate in this study: *

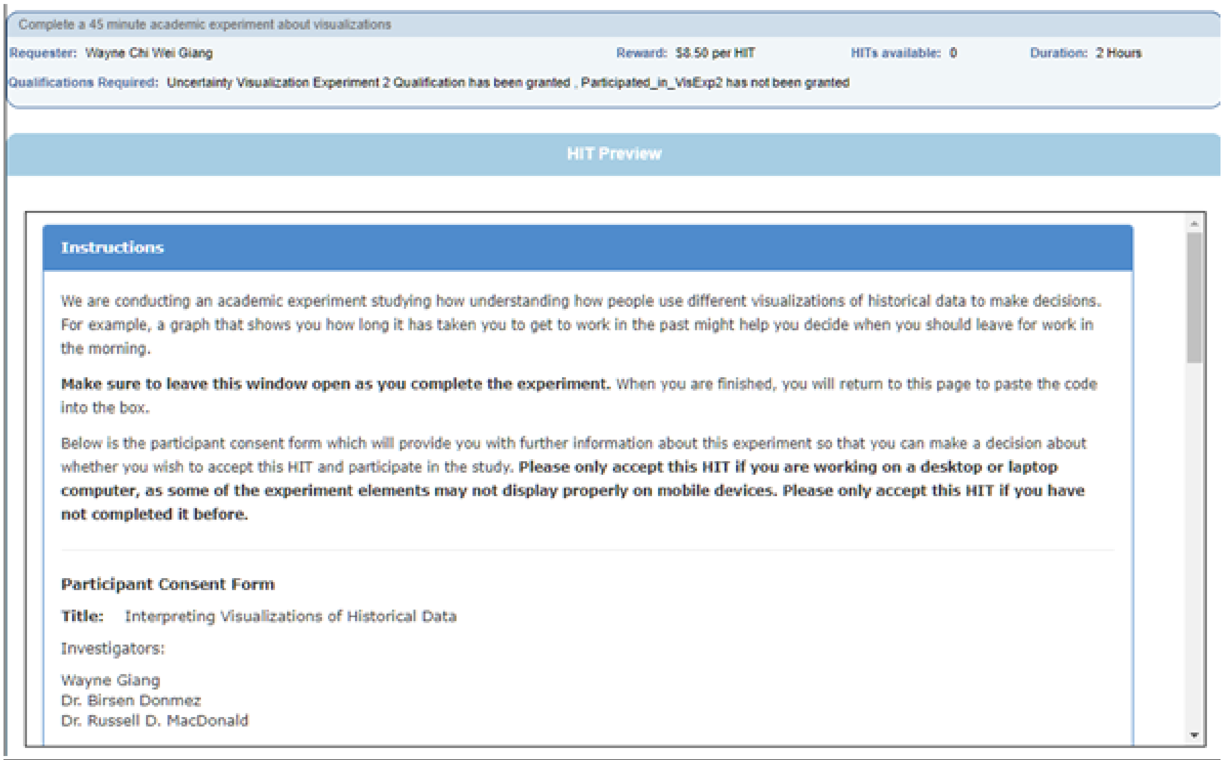
- ☐ I have read this consent form and I understand the research and what is expect of me.
- ☐ I understand that I am free to withdraw before or anytime during the study without the need to give any explanation by closing the browser window.
- ☐ I agree to participate in this study.

If I do not wish to participate in the research, I can just close this Internet browser window without agreeing to the above.

Next

0%

Amazon Mechanical Turk HIT for Experiment



Complete a 45 minute academic experiment about visualizations

Requester: Wayne Chi Wei Giang | Reward: $8.50 per HIT | HITs available: 0 | Duration: 2 Hours

Qualifications Required: Uncertainty Visualization Experiment 2 Qualification has been granted , Participated_in_VisExp2 has not been granted

HIT Preview

Instructions

We are conducting an academic experiment studying how understanding how people use different visualizations of historical data to make decisions. For example, a graph that shows you how long it has taken you to get to work in the past might help you decide when you should leave for work in the morning.

Make sure to leave this window open as you complete the experiment. When you are finished, you will return to this page to paste the code into the box.

Below is the participant consent form which will provide you with further information about this experiment so that you can make a decision about whether you wish to accept this HIT and participate in the study. **Please only accept this HIT if you are working on a desktop or laptop computer, as some of the experiment elements may not display properly on mobile devices. Please only accept this HIT if you have not completed it before.**

Participant Consent Form

Title: Interpreting Visualizations of Historical Data

Investigators:

Wayne Giang
Dr. Birsen Donmez
Dr. Russell D. MacDonald

Consent Form for Experiment

Participant Consent Form

Title: Interpreting Visualizations of Historical Data

Investigators:

Wayne Giang
Dr. Birsen Donmez
Dr. Russell D. MacDonald

Thank you for your interest in this research project. In order to decide whether you wish to participate in this research study, you should understand enough about the experimental procedures, and the risks and benefits of participating to be able to make an informed decision. This is known as the informed consent process.

If at any point you feel that the following details are unclear, or if you have any other questions or concerns, please feel free to contact me using the contact information provided above. If you decide that you would like to participate, please check off all checkboxes at the bottom of the page and enter your MTurk Worker ID. If you do not wish to participate, there is no need to complete the form and you may close this window and return the HIT.

Purpose

We are interested in understanding how people use different visualizations of historical data to make decisions. For example, a graph that shows you how long it has taken you to get to work in the past might help you decide when you should leave for work in the morning.

As a participant you will be asked to:
• Complete an experimental task where you make decisions based on historical data presented using different visualizations.
• Complete a questionnaire about the strategies you used during the experiment and about personality factors related to decision-making.

Procedure

There are two parts to this study. In the first part, you will take the role of a medical dispatcher who is responsible for estimating how long it will take to transfer patients between hospitals. You will be shown historical data about the length of similar patient transfers from the past, and information about the current patient transfer to help you with your estimates. In the second part of the study, you will fill out a questionnaire about the strategies you used while making these patient transfer time estimates as well as two questionnaires that examine personality factors that may be related to how you would use data in decision-making.

Risks

There are no major risks involved with this experiment, the tasks are not physiologically demanding, or psychologically stressing.

Benefits

There are several benefits to conducting this study. The most important benefit is your contribution to research in information visualization and decision support systems, which will guide the development of future decision support systems for expert decision makers such as medical dispatchers.

Compensation

This experiment will take approximately 45 minutes. You will receive $8.50 upon completion of the experiment.

Confidentiality

All information obtained during the study will be held in strict confidence. You will be identified with a study number only, and this study number will only be identifiable by the primary investigator. No names or identifying information will be used in any publication or presentation. No information identifying you will be transferred outside the investigators in this study.

Participation

Your participation in this study is voluntary. You can choose to not participate or withdraw at any time by closing this window before completing all the questions and returning the HIT. If you choose to withdraw, your data from this experiment will be deleted. If you choose to withdraw you will not be compensated for this study.

Questions

If you have any general questions about this study, please call 416.978.0881 or email wayne.giang [at] mail.utoronto.ca. You can also contact the Office of Research Ethics at ethics.review [at] utoronto.ca or 416-946-3273, if you have questions about your rights as participants.

To be completed by participants:

Please enter your MTurk Worker ID.
☐ I have read this consent form and I understand the research and what is expected of me.
☐ I understand that I am free to withdraw before or anytime during the study without the need to give any explanation by closing the browser window.
☐ I agree to participate in this study.

MTurk WorkerID: [ ]

Submit

Datasets used for experiment in Chapter 7

Symmetric Distributions

Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6
15.5	19.5	23.5	10.5	14.5	18.5
16	20	24	12.5	16.5	20.5
16.5	20.5	24.5	14	18	22
17	21	25	14.5	18.5	22.5
17.5	21.5	25.5	15.5	19.5	23.5
17.5	21.5	25.5	16	20	24
17.5	21.5	25.5	16.5	20.5	24.5
18	22	26	17	21	25
18	22	26	17	21	25
18	22	26	17.5	21.5	25.5
18.5	22.5	26.5	18	22	26
18.5	22.5	26.5	18.5	22.5	26.5
18.5	22.5	26.5	18.5	22.5	26.5
19	23	27	19	23	27
19	23	27	19	23	27
19	23	27	19.5	23.5	27.5
19	23	27	20	24	28
19	23	27	20	24	28
19.5	23.5	27.5	20.5	24.5	28.5
19.5	23.5	27.5	20.5	24.5	28.5
19.5	23.5	27.5	21	25	29
19.5	23.5	27.5	21	25	29
19.5	23.5	27.5	21.5	25.5	29.5
20	24	28	21.5	25.5	29.5
20	24	28	22	26	30
20	24	28	22	26	30
20	24	28	22.5	26.5	30.5
20.5	24.5	28.5	22.5	26.5	30.5
20.5	24.5	28.5	23	27	31
20.5	24.5	28.5	23	27	31
20.5	24.5	28.5	23.5	27.5	31.5
20.5	24.5	28.5	23.5	27.5	31.5
21	25	29	24	28	32
21	25	29	24	28	32
21	25	29	24.5	28.5	32.5
21	25	29	25	29	33
21	25	29	25	29	33
21.5	25.5	29.5	25.5	29.5	33.5
21.5	25.5	29.5	25.5	29.5	33.5
21.5	25.5	29.5	26	30	34

22	26	30	26.5	30.5	34.5
22	26	30	27	31	35
22	26	30	27	31	35
22.5	26.5	30.5	27.5	31.5	35.5
22.5	26.5	30.5	28	32	36
22.5	26.5	30.5	28.5	32.5	36.5
23	27	31	29.5	33.5	37.5
23.5	27.5	31.5	30	34	38
24	28	32	31.5	35.5	39.5
24.5	28.5	32.5	33.5	37.5	41.5

Skewed Distributions

Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6
15	19	23	12	16	20
15.5	19.5	23.5	13.5	17.5	21.5
15.5	19.5	23.5	14	18	22
16	20	24	14.5	18.5	22.5
16	20	24	15	19	23
16	20	24	15	19	23
16	20	24	15.5	19.5	23.5
16.5	20.5	24.5	15.5	19.5	23.5
16.5	20.5	24.5	16	20	24
16.5	20.5	24.5	16	20	24
16.5	20.5	24.5	16.5	20.5	24.5
16.5	20.5	24.5	16.5	20.5	24.5
16.5	20.5	24.5	17	21	25
17	21	25	17	21	25
17	21	25	17	21	25
17	21	25	17.5	21.5	25.5
17	21	25	17.5	21.5	25.5
17	21	25	17.5	21.5	25.5
17	21	25	18	22	26
17	21	25	18	22	26
17.5	21.5	25.5	18	22	26
17.5	21.5	25.5	18.5	22.5	26.5
17.5	21.5	25.5	18.5	22.5	26.5
17.5	21.5	25.5	19	23	27
17.5	21.5	25.5	19	23	27
17.5	21.5	25.5	19	23	27
17.5	21.5	25.5	19.5	23.5	27.5
18	22	26	19.5	23.5	27.5
18	22	26	19.5	23.5	27.5
18	22	26	20	24	28

18	22	26	20	24	28
18	22	26	20.5	24.5	28.5
18	22	26	20.5	24.5	28.5
18.5	22.5	26.5	21	25	29
18.5	22.5	26.5	21	25	29
18.5	22.5	26.5	21.5	25.5	29.5
18.5	22.5	26.5	21.5	25.5	29.5
19	23	27	22	26	30
19	23	27	22.5	26.5	30.5
19	23	27	22.5	26.5	30.5
19.5	23.5	27.5	23	27	31
19.5	23.5	27.5	23.5	27.5	31.5
19.5	23.5	27.5	24	28	32
20	24	28	25	29	33
20	24	28	25.5	29.5	33.5
20.5	24.5	28.5	26.5	30.5	34.5
21	25	29	27.5	31.5	35.5
22	26	30	29.5	33.5	37.5
23	27	31	32.5	36.5	40.5
26	30	34	39.5	43.5	47.5

Appendix G – Statistical Models for Chapter 7

Predictions on the Salient Central Tendency Point

Full Model

```
proc genmod data = Sym_exp.Data;
CLASS PartID Vis_type scenario_sd context;
model deviation = Vis_type|scenario_sd|context / link = logit dist = binomial type3 wald;
repeated subject = PartID;
run; quit;
```

Wald Statistics For Type 3 GEE Analysis			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
Vis_type	1	0.05	0.8235
scenario_sd	1	0.02	0.8893
Vis_type*scenario_sd	1	0.85	0.3578
context	2	143.42	<.0001
Vis_type*context	2	1.62	0.4454
scenario_sd*context	2	1.05	0.5923
Vis_ty*scenar*contex	2	1.03	0.5974

Final Model

```
proc genmod data = Sym_exp.Data;
CLASS PartID Vis_type scenario_sd context;
model deviation = context/ link = logit dist = binomial type3 wald;
repeated subject = PartID;
estimate 'No Context' intercept 1 Context 1   0 0 /exp;
estimate 'Likelihood' intercept 1 Context 0    1 0 /exp;
estimate 'Consequence' intercept 1 Context 0 0          1 /exp;
estimate 'No Context vs Consequence' Context 1        0 -1 /exp;
estimate 'No Context vs Likelihood' Context 1 -1 0/exp;
estimate 'Consequence vs Likelihood' Context 0        -1 1 /exp;
run; quit;
```

Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		-0.4638	0.1789	-0.8145	-0.1131	-2.59	0.0095
context	Baseline	1.3846	0.1829	1.0261	1.7430	7.57	<.0001
context	Likelihood	-3.1868	0.4734	-4.1147	-2.2589	-6.73	<.0001
context	Value	0.0000	0.0000	0.0000	0.0000	.	.

Wald Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
context	2	134.83	<.0001

Contrast Estimate Results									
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
No Context	0.7152	0.6287	0.7883	0.9207	0.2010	0.05	0.5267	1.3148	20.97	<.0001
Exp(No Context)				2.5111	0.5049	0.05	1.6933	3.7239		
Likelihood	0.0253	0.0111	0.0568	-3.6507	0.4290	0.05	-4.4915	-2.8098	72.41	<.0001
Exp(Likelihood)				0.0260	0.0111	0.05	0.0112	0.0602		
Consequence	0.3861	0.3069	0.4717	-0.4638	0.1789	0.05	-0.8145	-0.1131	6.72	0.0095
Exp(Consequence)				0.6289	0.1125	0.05	0.4428	0.8930		

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
No Context vs Consequence	0.7997	0.7362	0.8511	1.3846	0.1829	0.05	1.0261	1.7430	57.32	<.0001
Exp(No Context vs Consequence)				3.9931	0.7302	0.05	2.7902	5.7144		
No Context vs Likelihood	0.9898	0.9749	0.9959	4.5714	0.4648	0.05	3.6604	5.4824	96.73	<.0001
Exp(No Context vs Likelihood)				96.6778	44.9359	0.05	38.8765	240.4172		
Consequence vs Likelihood	0.9603	0.9054	0.9839	3.1868	0.4734	0.05	2.2589	4.1147	45.31	<.0001
Exp(Consequence vs Likelihood)				24.2113	11.4620	0.05	9.5730	61.2334		

<u>Direction of Predictions relative to Central Tendency Point</u>

Full Model

```
proc genmod data = Sym_exp.Data_dev;
CLASS PartID Vis_type scenario_sd context;
model AboveCT = Vis_type|context|scenario_sd/ link = logit dist = binomial type3 wald;
repeated subject = PartID;
run; quit;
```

Wald Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
Vis_type	1	3.67	0.0555
context	2	42.58	<.0001
Vis_type*context	2	1.83	0.4005
scenario_sd	1	4.35	0.0370

Wald Statistics For Type 3 GEE Analysis			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
Vis_type*scenario_sd	1	0.17	0.6803
scenario_sd*context	2	1.92	0.3835
Vis_ty*scenar*contex	2	4.68	0.0964

Final Model

```
proc genmod data = Sym_exp.Data_dev descending;
CLASS PartID Vis_type scenario_sd context;
model AboveCT = context scenario_sd / link = logit dist = binomial type3 wald;
repeated subject = PartID;
estimate 'No Context' intercept 1 Context 1   0 0 /exp;
estimate 'Likelihood' intercept 1 Context 0    1 0 /exp;
estimate 'Consequence' intercept 1 Context 0 0           1 /exp;
estimate 'No Context vs Consequence' Context 1        0 -1 /exp;
estimate 'Likelihood vs No Context' Context -1 1 0/exp;
estimate 'Likelihood vs Consequence' Context 0        1 -1 /exp;
estimate 'Smaller SD' intercept 1 scenario_sd 1          0 /exp;
estimate 'Larger SD' intercept 1 Context 0    1 /exp;
estimate 'Smaller vs Larger SD' scenario_sd 1            -1 /exp;
run; quit;
```

Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter		**Estimate**	**Standard Error**	**95% Confidence Limits**		**Z**	**Pr > |Z|**
Intercept		0.4300	0.2250	-0.0110	0.8709	1.91	0.0560
context	Baseline	-0.0367	0.2911	-0.6072	0.5337	-0.13	0.8996
context	Likelihood	2.0386	0.3190	1.4135	2.6637	6.39	<.0001
context	Value	0.0000	0.0000	0.0000	0.0000	.	.
scenario_sd	2	-0.5650	0.2328	-1.0213	-0.1087	-2.43	0.0152
scenario_sd	5	0.0000	0.0000	0.0000	0.0000	.	.

Wald Statistics For Type 3 GEE Analysis			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
context	2	43.35	<.0001
scenario_sd	1	5.89	0.0152

Contrast Estimate Results

Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
No Context	0.5277	0.3713	0.6787	0.1107	0.3251	0.05	-0.5265	0.7480	0.12	0.7334
Exp(No Context)				1.1171	0.3632	0.05	0.5907	2.1127		
Likelihood	0.8990	0.8411	0.9374	2.1861	0.2651	0.05	1.6665	2.7057	68.00	<.0001
Exp(Likelihood)				8.9004	2.3595	0.05	5.2936	14.9645		
Consequence	0.5368	0.4429	0.6282	0.1475	0.1923	0.05	-0.2295	0.5245	0.59	0.4432
Exp(Consequence)				1.1589	0.2229	0.05	0.7949	1.6896		
No Context vs Consequence	0.4908	0.3527	0.6304	-0.0367	0.2911	0.05	-0.6072	0.5337	0.02	0.8996
Exp(No Context vs Consequence)				0.9639	0.2806	0.05	0.5449	1.7053		
Likelihood vs No Context	0.8885	0.7905	0.9439	2.0753	0.3813	0.05	1.3280	2.8227	29.63	<.0001
Exp(Likelihood vs No Context)				7.9673	3.0378	0.05	3.7736	16.8214		

Contrast Estimate Results										
		Mean					L'Beta			
Label	Mean Estimate	Confidence Limits		L'Beta Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > Chi Sq
Likelihood vs Consequence	0.8848	0.8043	0.9349	2.0386	0.3190	0.05	1.4135	2.6637	40.85	<.0001
Exp(Likelihood vs Consequence)				7.6799	2.4495	0.05	4.1102	14.3499		
Smaller SD	0.6300	0.5243	0.7246	0.5323	0.2219	0.05	0.0973	0.9673	5.75	0.0165
Exp(Smaller SD)				1.7028	0.3779	0.05	1.1021	2.6308		
Larger SD	0.8990	0.8411	0.9374	2.1861	0.2651	0.05	1.6665	2.7057	68.00	<.0001
Exp(Larger SD)				8.9004	2.3595	0.05	5.2936	14.9645		
Smaller vs Larger SD	0.3624	0.2648	0.4729	-0.5650	0.2328	0.05	-1.0213	-0.1087	5.89	0.0152
Exp(Smaller vs Larger SD)				0.5684	0.1323	0.05	0.3601	0.8970		

Distance between Prediction and Central Tendency Point

Full Model

```
proc genmod data = Sym_exp.Data_dev descending;
CLASS PartID Vis_type scenario_sd context Dev_Cut;
model Dev_Cut = Vis_type|context|scenario_sd/ link = clogit dist = mult type3 wald ;
repeated subject = PartID;
run; quit;
```

Wald Statistics For Type 3 GEE Analysis			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
Vis_type	1	167.76	<.0001
context	2	366.56	<.0001
Vis_type*context	2	261.19	<.0001
scenario_sd	1	232.64	<.0001
Vis_type*scenario_sd	1	565.64	<.0001
scenario_sd*context	2	102.64	<.0001
Vis_ty*scenar*contex	1	3.96	0.0465

Final Model

```
proc genmod data = Sym_exp.Data_dev descending;
CLASS PartID Vis_type scenario_sd context story2 Dev_Cut;
model Dev_Cut = scenario_sd|context/ link = clogit dist = mult type3 wald ;
repeated subject = PartID;

estimate 'For Smaller SD: No Context vs. Consequence' Context 1 0 -1 scenario_sd*Context 1 0 -1 0 0 0/exp;
estimate 'For Smaller SD: No Context vs. Likelihood' Context 1 -1 0 scenario_sd*Context 1 -1 0 0 0 0/exp;
estimate 'For Smaller SD: Consequence vs. Likelihood' Context 0 -1 1 scenario_sd*Context 0 -1 1 0 0 0/exp;

estimate 'For Larger SD: No Context vs. Consequence' Context 1 0 -1 scenario_sd*Context 0 0 0 1 0 -1/exp;
estimate 'For Larger SD: No Context vs. Likelihood' Context 1 -1 0 scenario_sd*Context 0 0 0 1 -1 0/exp;
estimate 'For Larger SD: Consequence vs. Likelihood' Context 0 -1 1 scenario_sd*Context 0 0 0 0 -1 1/exp;

estimate 'For No Context: Smaller vs Larger SD' scenario_sd 1 -1 scenario_sd*Context 1 0 0 -1 0 0/exp;
estimate 'For Consequence: Smaller vs Larger SD' scenario_sd 1 -1 scenario_sd*Context 0 0 1 0 0 -1/exp;
estimate 'For Likelihood: Smaller vs Larger SD' scenario_sd 1 -1 scenario_sd*Context 0 1 0 0 -1 0/exp;
run; quit;
```

Analysis Of GEE Parameter Estimates								
Empirical Standard Error Estimates								
Parameter			**Estimate**	**Standard Error**	**95% Confidence Limits**		**Z**	**Pr > |Z|**
Intercept1			-2.5918	0.4190	-3.4131	-1.7706	-6.19	<.0001
Intercept2			-0.9343	0.2950	-1.5126	-0.3561	-3.17	0.0015

Analysis Of GEE Parameter Estimates

Empirical Standard Error Estimates

Parameter			Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept3			1.0018	0.2686	0.4752	1.5283	3.73	0.0002
scenario_sd	2		-1.1012	0.3487	-1.7847	-0.4178	-3.16	0.0016
scenario_sd	5		0.0000	0.0000	0.0000	0.0000	.	.
context	Baseline		-2.0076	0.4484	-2.8864	-1.1288	-4.48	<.0001
context	Likelihood		3.6646	0.4708	2.7419	4.5873	7.78	<.0001
context	Value		0.0000	0.0000	0.0000	0.0000	.	.
scenario_sd*context	2	Baseline	-1.5855	1.1733	-3.8850	0.7141	-1.35	0.1766
scenario_sd*context	2	Likelihood	-2.5288	0.4354	-3.3823	-1.6754	-5.81	<.0001
scenario_sd*context	2	Value	0.0000	0.0000	0.0000	0.0000	.	.
scenario_sd*context	5	Baseline	0.0000	0.0000	0.0000	0.0000	.	.
scenario_sd*context	5	Likelihood	0.0000	0.0000	0.0000	0.0000	.	.
scenario_sd*context	5	Value	0.0000	0.0000	0.0000	0.0000	.	.

Wald Statistics For Type 3 GEE Analysis

Source	DF	Chi-Square	Pr > ChiSq
scenario_sd	1	43.45	<.0001
context	2	116.71	<.0001
scenario_sd*context	2	33.77	<.0001

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
For Smaller SD: No Context vs. Consequence	0.0268	0.0036	0.1715	-3.5931	1.0296	0.05	-5.6111	-1.5750	12.18	0.0005
Exp(For Smaller SD: No Context vs. Consequence)				0.0275	0.0283	0.05	0.0037	0.2070		
For Smaller SD: No Context vs. Likelihood	0.0088	0.0013	0.0586	-4.7288	0.9960	0.05	-6.6810	-2.7767	22.54	<.0001
Exp(For Smaller SD: No Context vs. Likelihood)				0.0088	0.0088	0.05	0.0013	0.0622		
For Smaller SD: Consequence vs. Likelihood	0.2431	0.1468	0.3748	-1.1358	0.3184	0.05	-1.7599	-0.5116	12.72	0.0004
Exp(For Smaller SD: Consequence vs. Likelihood)				0.3212	0.1023	0.05	0.1721	0.5995		
For Larger SD: No Context vs. Consequence	0.1184	0.0528	0.2444	-2.0076	0.4484	0.05	-2.8864	-1.1288	20.05	<.0001
Exp(For Larger SD: No Context				0.1343	0.0602	0.05	0.0558	0.3234		

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
vs. Consequence)										
For Larger SD: No Context vs. Likelihood	0.0034	0.0013	0.0090	-5.6722	0.4944	0.05	-6.6412	-4.7032	131.64	<.0001
Exp(For Larger SD: No Context vs. Likelihood)				0.0034	0.0017	0.05	0.0013	0.0091		
For Larger SD: Consequence vs. Likelihood	0.0250	0.0101	0.0605	-3.6646	0.4708	0.05	-4.5873	-2.7419	60.59	<.0001
Exp(For Larger SD: Consequence vs. Likelihood)				0.0256	0.0121	0.05	0.0102	0.0644		
For No Context: Smaller vs Larger SD	0.0638	0.0079	0.3677	-2.6867	1.0942	0.05	-4.8314	-0.5420	6.03	0.0141
Exp(For No Context: Smaller vs Larger SD)				0.0681	0.0745	0.05	0.0080	0.5816		
For Consequence: Smaller vs Larger SD	0.2495	0.1437	0.3970	-1.1012	0.3487	0.05	-1.7847	-0.4178	9.97	0.0016
Exp(For Consequence				0.3325	0.1159	0.05	0.1678	0.6585		

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
e: Smaller vs Larger SD)										
For Likelihood: Smaller vs Larger SD	0.0258	0.0150	0.0441	-3.6301	0.2823	0.05	-4.1833	-3.0769	165.40	<.0001
Exp(For Likelihood: Smaller vs Larger SD)				0.0265	0.0075	0.05	0.0152	0.0461		

<u>Confidence in Predictions</u>

Full Model

proc mixed data = Sym_exp.Data;
CLASS PartID Vis_type scenario_sd context story2;
model Est_conf = Vis_type|scenario_sd|context/ ddfm=satterth RESIDUAL solution outp = res;
run; quit;

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Vis_type	1	936	5.14	0.0236
scenario_sd	1	936	4.13	0.0425
Vis_type*scenario_sd	1	936	2.87	0.0908
context	2	936	0.14	0.8715
Vis_type*context	2	936	0.28	0.7589
scenario_sd*context	2	936	0.05	0.9509
Vis_ty*scenar*contex	2	936	0.43	0.6534

Final Model

```
proc mixed data = Sym_exp.Data;
CLASS PartID Vis_type scenario_sd context story2;
model Est_conf = Vis_type|scenario_sd context/ ddfm=satterth RESIDUAL solution outp = res;
repeated / subject = PartID type = cs;
estimate 'Smaller SD Median' intercept 1 Vis_type 1 0 scenario_sd 1 0 Vis_type*scenario_sd 1 0 0 0 /cl;
estimate 'Larger SD Median' intercept 1 Vis_type 1 0 scenario_sd 0 1  Vis_type*scenario_sd 0 1 0 0 /cl;
estimate 'For Smaller SD: Median vs. Boxplot' Vis_type 1 -1 Vis_type*scenario_sd 1 0 -1 0 /cl;
estimate 'For Larger SD: Median vs. Boxplot' Vis_type 1 -1 Vis_type*scenario_sd 0 1 0 -1 /cl;
estimate 'For Median-only: Smaller SD vs. Larger SD' scenario_sd 1 -1 Vis_type*scenario_sd 1 -1 0 0 /cl;
estimate 'For Boxplot: Smaller vs. Larger SD' scenario_sd 1 -1 Vis_type*scenario_sd 0 0 1 -1 /cl;
estimate 'Smaller SD Median vs. Larger SD Boxplot' Vis_type 1 -1 scenario_sd 1 -1 Vis_type*scenario_sd 1 0 0 -1 /
cl;
run; quit;
```

Solution for Fixed Effects

Effect	Vis_type	context	scenario_sd	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept				53.7764	2.2819	113	23.57	<.0001
Vis_type	Central_Tendency			5.7342	1.1991	864	4.78	<.0001
Vis_type	Quantiles			0	.	.	.	.
scenario_sd			2	5.3924	1.1991	864	4.50	<.0001
scenario_sd			5	0	.	.	.	.
Vis_type*scenario_sd	Central_Tendency		2	-4.9030	1.6957	864	-2.89	0.0039
Vis_type*scenario_sd	Central_Tendency		5	0	.	.	.	.
Vis_type*scenario_sd	Quantiles		2	0	.	.	.	.
Vis_type*scenario_sd	Quantiles		5	0	.	.	.	.
context		Baseline		-0.1424	1.0384	864	-0.14	0.8910

Solution for Fixed Effects								
Effect	**Vis_type**	**context**	**scenario_sd**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
context		Likelihood		0.7247	1.0384	864	0.70	0.4854
context		Value		0	.	.	.	.

Type 3 Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Vis_type	1	864	14.99	0.0001
scenario_sd	1	864	12.03	0.0005
Vis_type*scenario_sd	1	864	8.36	0.0039
context	2	864	0.40	0.6698

Estimates								
Label	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Smaller SD Median	60.1941	2.2017	98.6	27.34	<.0001	0.05	55.8252	64.5630
Larger SD Median	59.7046	2.2017	98.6	27.12	<.0001	0.05	55.3358	64.0735
For Smaller SD: Median vs. Boxplot	0.8312	1.1991	864	0.69	0.4884	0.05	-1.5222	3.1847
For Larger SD: Median vs. Boxplot	5.7342	1.1991	864	4.78	<.0001	0.05	3.3807	8.0876
For Median-only: Smaller SD vs. Larger SD	0.4895	1.1991	864	0.41	0.6832	0.05	-1.8640	2.8429
For Boxplot: Smaller vs. Larger SD	5.3924	1.1991	864	4.50	<.0001	0.05	3.0390	7.7458
Smaller SD Median vs. Larger SD Boxplot	6.2236	1.1991	864	5.19	<.0001	0.05	3.8702	8.5771

Appendix H – Prediction Strategy Tables for Chapter 7

Table 8: Percentage of participants who rated their predictions to be an Optimistic (O), Average (A), or Pessimistic (P) case for each of the 10 prediction strategies for symmetric historical data distributions

No Context		**Other** **Median-only:** O: 15% A:68% P: 19% **Boxplot:** O: 7% A: 81% P: 11%		
		Prediction Location		
		Below	**Median**	**Above**
Prediction Probability	**>50%**	**Median-only:** O: 60% A:40% P: 0% **Boxplot:** O: 20% A: 80% P: 0%	**Median-only:** O: 0% A:100% P: 0% **Boxplot:** O: 0 A: 100% P: 0%	**Median-only:** O: 0% A:29% P: 71% **Boxplot:** O: 0% A: 29% P: 71%
	50%		**Median-only:** O: 0% A:100% P: 0% **Boxplot:** O: 3% A: 97% P: 0%	
	<50%		**Median-only:** O: 0% A:100% P: 0% **Boxplot:** O: 0% A: 100% P: 0%	
Consequence		**Other** **Median-only:** O: 53% A:33% P: 13% **Boxplot:** O: 56% A: 31% P: 13%		
		Prediction Location		
		Below	**Median**	**Above**
Prediction Probability	**>50%**	**Median-only:** O: 83% A:17% P: 0% **Boxplot:** O: 50% A: 33% P: 17%	**Median-only:** O: 26% A:62% P: 12% **Boxplot:** O: 25 A: 50% P: 25%	**Median-only:** O: 0% A:33% P: 67% **Boxplot:** O: 11% A: 11% P: 78%
	50%		**Median-only:** O: 0% A:88% P: 12% **Boxplot:** O: 3% A: 75% P: 25%	
	<50%	**Median-only:** O: 83% A:17% P: 0% **Boxplot:** O: 50% A: 33% P: 17%	**Median-only:** O: 0% A:100% P: 0% **Boxplot:** O: 0% A: 100% P: 0%	
Likelihood		**Other** **Median-only:** O: 26% A:41% P: 33% **Boxplot:** O: 19% A: 56% P: 26%		
		Prediction Location		
		Below	**Median**	**Above**
Predicti on	**>50%**			**Median-only:** O: 22% A:33% P: 44% **Boxplot:** O: 22% A: 25% P: 53%

	50%	**Median-only:** O: 0% A:100% P: 0% **Boxplot:** O: 0% A: 100% P: 0%		**Median-only:** O: 0% A:100% P: 0% **Boxplot:** O: 0% A: 100% P: 0%
	<50%			**Median-only:** O: 8% A:67% P: 25% **Boxplot:** O: 0% A: 58% P: 42%

Table 9: Percentage of participants who rated their predictions to be an Optimistic (O), Average (A), or Pessimistic (P) case for each of the 10 prediction strategies for the Median-only visualization for right-skewed historical data distributions

No Context		**Other** **Median-only:** O: 21% A: 55% P: 24%		
		Prediction Location		
		Below	**Median**	**Above**
Prediction Probability	**>50%**	**Median-only:** O: 0% A: 0% P: 100%	**Median-only:** O: 0% A: 90% P: 10%	**Median-only:** O: 50% A: 0% P: 50%
	50%		**Median-only:** O: 0% A: 96% P: 4%	**Median-only:** O: 0% A: 100% P: 0%
	<50%	**Median-only:** O: 0% A: 100% P: 0%	**Median-only:** O: 0% A: 100% P: 0%	
Consequence		**Other** **Median-only:** O: 36% A: 30% P: 34%		
		Prediction Location		
		Below	**Median**	**Above**
Prediction Probability	**>50%**	**Median-only:** O: 33% A: 17% P: 50%	**Median-only:** O: 25 A: 50% P: 25%	**Median-only:** O: 17% A: 50% P: 33%
	50%	**Median-only:** O: 0% A: 100% P: 0%	**Median-only:** O: 17% A: 67% P: 17%	**Median-only:** O: 0% A: 100% P: 0%
	<50%		**Median-only:** O: 0% A: 100% P: 0%	**Median-only:** O: 50% A: 50% P: 0%
Likelihood		**Other** **Median-only:** O: 19% A: 50% P: 31%		
		Prediction Location		
		Below	**Median**	**Above**

Prediction Probability	>50%	**Median-only:** O: 0% A: 100% P: 0%	**Median-only:** O: 0% A: 100% P: 0%	**Median-only:** O: 7% A: 32% P: 61%
	50%		**Median-only:** O: 0% A: 100% P: 0%	**Median-only:** O: 0% A: 67% P: 33%
	<50%	**Median-only:** O: 0% A: 100% P: 0%		**Median-only:** O: 25% A: 38% P: 38%

Table 10: Percentage of participants who rated their predictions to be an Optimistic (O), Average (A), or Pessimistic (P) case for each of the 10 prediction strategies for the Boxplot visualization for right-skewed historical data distributions

No Context		**Other** **Boxplot:** O: 17% A: 67% P: 17%		
		Prediction Location		
		Below	**Median**	**Above**
Prediction Probability	**>50%**		**Boxplot:** O: 0% A: 100% P: 0%	**Boxplot:** O: 6% A: 65% P: 29%
	50%		**Boxplot:** O: 6% A: 82% P: 12%	**Boxplot:** O: 0% A: 0% P: 100%
	<50%	**Boxplot:** O: 0% A: 100% P: 0%	**Boxplot:** O: 0% A: 100% P: 0%	**Boxplot:** O: 0% A: 100% P: 0%
Consequence		**Other** **Boxplot:** O: 32% A: 41% P: 27%		
		Prediction Location		
		Below	**Median**	**Above**
Prediction Probability	**>50%**	**Boxplot:** O: 33% A: 17% P: 50%	**Boxplot:** O: 25 A: 50% P: 25%	**Boxplot:** O: 16% A: 37% P: 47%
	50%	**Boxplot:** O: 0% A: 100% P: 0%	**Boxplot:** O: 0% A: 100% P: 0%	**Boxplot:** O: 0% A: 100% P: 0%
	<50%	**Boxplot:** O: 100% A: 0% P: 0%	**Boxplot:** O: 0% A: 100% P: 0%	**Boxplot:** O: 100% A: 0% P: 0%
Likelihood		**Other** **Boxplot:** O: 24% A: 32% P: 44%		
		Prediction Location		
		Below	**Median**	**Above**
Prediction Probability	**>50%**	**Boxplot:** O: 0% A: 100% P: 0%	**Boxplot:** O: 0% A: 100% P: 0%	**Boxplot:** O: 8% A: 30% P: 62%
	50%			**Boxplot:** O: 17% A: 67% P: 17%
	<50%	**Boxplot:** O: 0% A: 100% P: 0%		**Boxplot:** O: 13% A: 62% P: 25%

Appendix I – Skewed Distribution Study (Chapter 7)

Methods

Experiment Design

Experiment 2 used the same experimental design as Experiment 1 (Table 3), with context information, visualization type, and variability magnitude as the independent variables, as well as the same experimental tasks. The major difference was the use of skewed distributions instead of symmetric distributions for the underlying historical data.

Six datasets were created from Burr Type XII distribution using R's actuar package, with shape parameters 0.74, 6.0, and a scale parameter of 37. Each of the datasets consisted of 50 observations that were evenly spaced throughout the distribution (e.g., 1st, 3rd, 5th... 99th percentile values of the distribution), but were then rounded to the nearest half minute. The Burr Type XII distribution has been used to model travel time distributions (Guessous, Aron, Bhouri, & Cohen, 2014), and the distribution parameters were fit to historical medical transport data in order to produce distribution shapes that were realistic for the task. The distribution was normalized and then transformed to create datasets with either a SD of 2 minutes or 5 minutes. The means of the distributions varied between 18 and 28. This procedure produced 6 datasets that were right skewed; with datasets with the smaller SD having skewness scores of 1.6 and kurtosis scores of 3.5, and datasets with the larger SD having skewness scores of 1.5 and kurtosis scores of 3.3.

Participants

An additional 110 potential participants were recruited using the paid screening questionnaire on MTurk, of which 86 qualified and were invited to participate in the study along with participants who had qualified from the previous screening questionnaire but had not participated in Experiment 1. The screening questionnaire took approximately 15 minutes and participants were compensated US$1.5 for filling out the screening questionnaire.

Eighty-six participants completed the experiment, but the data from seven participants were removed due to incomplete or abnormal response behavior during the experiment. The remaining 79 participants' data were used for analysis (50 male, 29 female, mean age = 35.2 years, SD =

9.1 years). Participants scored an average of 88% (SD = 10%) on the screening questionnaire. The majority of participants reported completing post-secondary education (n=55), with 36 participants taking at least one probability or statistics course at the post-secondary level, and 21 additional participants reported taking a course at the high school level. The experiment took 45 minutes and participants were compensated US$8.5.

Hypotheses

It was hypothesized that when participants are able to see the skewed underlying distributions (e.g., in the Boxplot condition), participants would be more likely to choose predictions that are away from the presented central tendency point, in the direction of the skewed distribution.

Results on Prediction Behavior

Predictions on the Salient Central Tendency Point

For predictions made using the Median-only visualization, where participants were not aware of the skewness of the underlying data, participants showed similar rates of using the salient central tendency point of the visualization (i.e., the median) as with the symmetric historical data in Experiment 1; 75% of predictions made with no-information, 34% of predictions made with consequence-information, and just 2% of predictions made using likelihood-information were on the saliently presented central tendency point of the visualizations (i.e. the median). For predictions made using the Boxplot visualization, 46% of predictions made with no-information, 23% of predictions made with consequence-information, and 3% of predictions made using the likelihood-information were on the salient central tendency point.

A logistic regression analysis found that the interaction between context and visualization was significant, $\chi^2(2)= 12.45$, p=.002, as was the interaction between visualization type and variability magnitude, $\chi^2(1)= 5.11$, p=.02, and the main effect of context, $\chi^2(2)= 87.93$, p<.0001. The regression model confirmed that when participants could see the skewed distributions in the Boxplot visualization condition, they were less likely to choose the central tendency point as their prediction, and this affect occurred across all three context conditions. Furthermore, it appeared that the influence of the skewed distributions was strongest in the no-information condition.

Direction of Predictions relative to the Salient Central Tendency Point

For trials where the predictions deviated from the salient central tendency point, 78% of the predictions were above the median (Figure 26). A logistic regression analysis found that the main effects of context, $\chi^2(2)= 32.1$, p<.0001, visualization type, $\chi^2(1)= 8.8$, p=.003, and variability magnitude, $\chi^2(1)= 11.7$, p=.0006, were significant. The results showed that the Median-only visualization resulted in fewer predictions above the central tendency point than the Boxplot visualization (OR: 0.5, 95% CI: 0.4-0.8), providing further evidence that the skewness and shape of the underlying distribution plays a role in how participants choose their prediction. Otherwise, the influence of variability magnitude and context were similar to those found in Experiment 1.

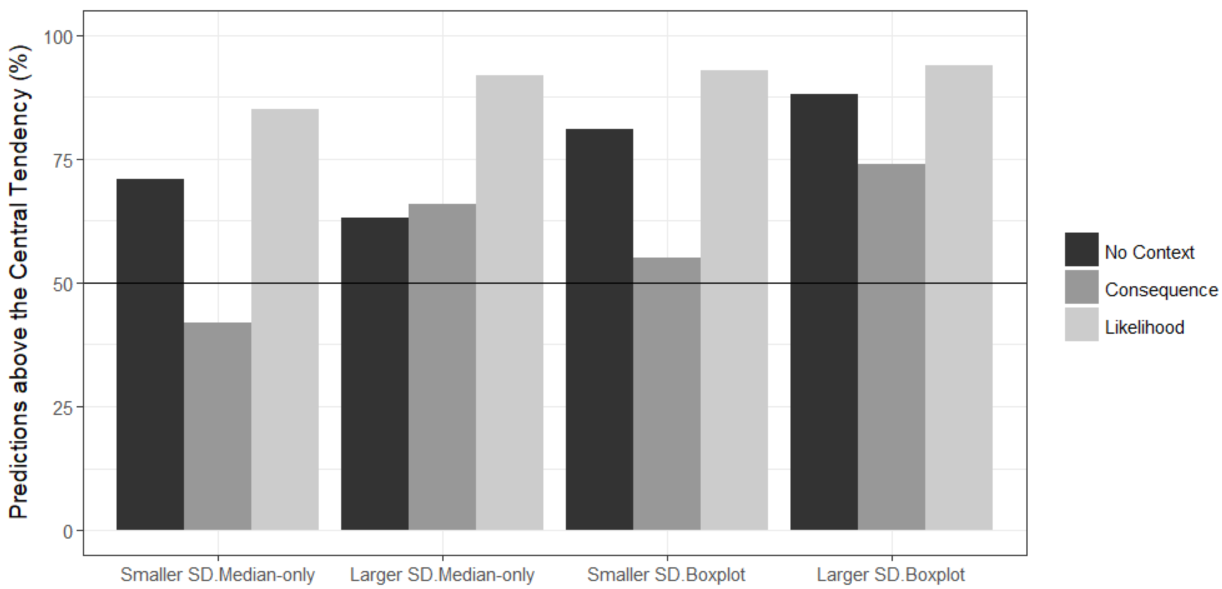


Figure 26: Percentage of trials with predictions above the salient central tendency point as opposed to below for right-skewed distributions

Distance between Prediction and the Salient Central Tendency Point

For predictions that were not on the salient central tendency point, the distance between the prediction and the salient central tendency point was divided into 4 levels with roughly equal numbers of observations: between 0.5 and 1 (n=242), between 1 and 2 (n=170), between 2 and 4 (n=94), and greater than 4 (n=135) minutes; the binning was performed because the data was highly non-normal. Figure 27 shows the number of predictions within each category across the different variability magnitude and context conditions.

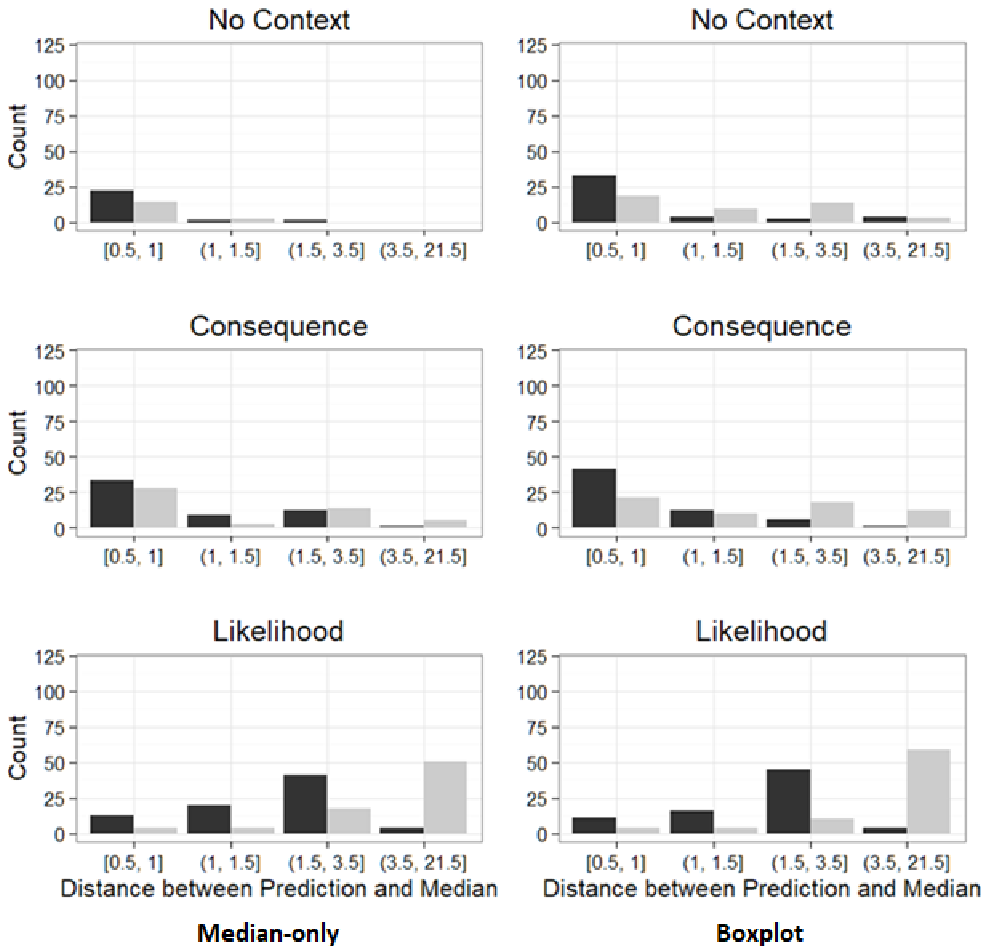


Figure 27: Number of predictions within each distance category for each level of variability magnitude and context information with the Median-only visualization (left) and the Boxplot visualization (right) for right-skewed distributions

An ordered logistic regression analysis found that the interaction between context and variability magnitude of the dataset was significant, $\chi^2(2)=20.7$, $p<.0001$, the interaction between visualization type and variability magnitude, $\chi^2(1)=14.2$, $p=.0002$, as were the main effects of visualization type, $\chi^2(1)=6.2$, $p=.01$, context information, $\chi^2(2)= 100.4$, $p<.0001$, and variability magnitude of the dataset, $\chi^2(1)= 82.9$, $p<.0001$. Trials with smaller SD tended to result in predictions closer to the median than trials with larger SD for the no-information (OR: 0.07, 95% CI: 0.01, 0.58), consequence-information (OR: 0.33, 95% CI: 0.17, 0.66), and likelihood-information (OR: 0.03, 95% CI: 0.02, 0.05) conditions. The skewed distributions resulted in differences between the Median-only and Boxplot visualizations which were not found in Experiment 1. For the larger SD condition, predictions made with the Boxplot visualization were more likely to be further away from the median than predictions made with the Median-only visualization. However, in the smaller SD condition, there was no difference between the two visualizations. As expected, within the two visualization conditions, predictions made with larger SD datasets were more likely to be further away from the median than predictions made with smaller SD datasets. Otherwise, results were similar to those found in Experiment 1.

Confidence in Predictions

A linear mixed analysis showed that only the main effects of visualization type, $\chi^2(1)= 11.3$, $p=.0008$, and variability magnitude, $\chi^2(1)= 21.8$, $p<.0001$, were significant. The Median-only visualization resulted in higher confidence than the Boxplot visualization ($\Delta= 3.1$, 95% CI: 1.3-4.9), and smaller SD datasets resulted in higher confidence than larger SD datasets ($\Delta= 4.3$, 95% CI: 2.5-6.1). Unlike in Experiment 1, the interaction between visualization type and variability magnitude was not significant, $\chi^2(1)= 2.9$, $p=.09$, however, the general trend of the results were similar to those found in Experiment 1.

Prediction Behavior

Overall, participants' prediction behaviors were similar to those found with the symmetric distributions in terms of the influence of contextual information and variability magnitude. However, it appeared that the skewed distributions had a large influence on prediction behavior between the two visualizations. Prediction behavior in the Median-only visualization appeared to be similar to prediction behavior with symmetric distributions. The Boxplot visualization, however, resulted in a greater number of predictions away from the central tendency as was hypothesized. Furthermore, the influence of the skewed distributions (i.e., the Boxplot visualization) appeared to be greatest in the no-information condition.

Results on Prediction Strategy

Probability Rating and Prediction Location

Figure 28 provides an overview of prediction strategies by presenting the location of participants' predictions relative to the salient central tendency point (i.e., the median) and their probability ratings. The figure illustrated the influence of the right-skewed distributions on predictions made with the Boxplot visualization. Within each of the three context conditions, each participant was assigned to nine different strategy categories based on these two dimensions: location with respect to central tendency (below, on, above) x probability rating (below 50%, 50%, above 50%). For a context condition, participants were assigned into one of these nine categories which represented the majority of their predictions. The breakdown is presented in Table 11. When there was no majority, the participant was assigned to a tenth category (labeled "Other" in Table 11).

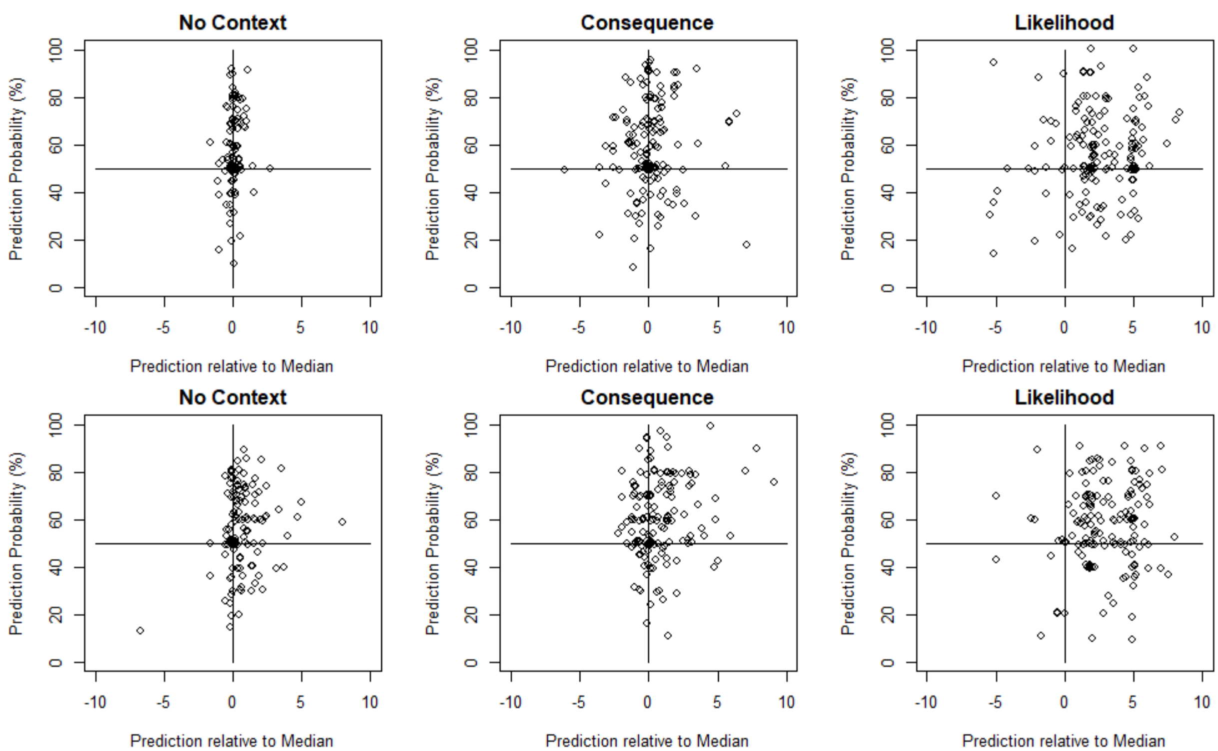


Figure 28: Predictions made by participants assessed across two dimensions: their location relative to the salient central tendency point and the prediction probability for right-skewed distributions

Table 11: Participant prediction strategies across the three context conditions. The strategy used by the plurality of participants is bolded.

Median-only		**No-context** Other = **33**			**Consequence** Other = **53**			**Likelihood** Other = **32**		
		Prediction Location								
		Below	*Median*	*Above*	*Below*	*Median*	*Above*	*Below*	*Median*	*Above*
Prediction Probability	*>50%*	1	10	2	4	4	6	1	0	28
	50%	0	25	1	2	6	0	0	1	6
	<50%	1	6	0	0	2	2	3	0	8

Boxplot		**No-context** Other = **30**			**Consequence** Other = **41**			**Likelihood** Other = 25		
		Prediction Location								
		Below	*Median*	*Above*	*Below*	*Median*	*Above*	*Below*	*Median*	*Above*
Prediction Probability	*>50%*	0	5	17	6	4	19	1	1	**37**
	50%	0	17	1	1	3	1	0	0	6
	<50%	1	4	4	1	1	2	1	0	8

As expected, the results provide evidence that the use of skewed-distributions with the Boxplot visualization shifted the prediction strategies used by participants in the direction of the skew, and this effect was strong for both the no-information and consequence-information conditions. In particular, the number of participants who used *Strategy-2* (i.e., Above Median and >50%) was much higher for the Boxplot than the Median-only visualization. The majority of the participants within this strategy category specified that their prediction was an average case (Table 12) in the post-experiment questionnaire for the no-information condition, providing evidence that participants used the skewness information to locate an alternative central tendency measure (e.g., the mean).

Table 12: The type of prediction indicated by the majority of participants within each strategy category: optimistic (Opt), average (Ave), or pessimistic (Pes).

Boxplot		No-context Other = Ave			Consequence Other = Opt			Likelihood Other = Ave		
		Prediction Location								
		Below	*Median*	*Above*	*Below*	*Median*	*Above*	*Below*	*Median*	*Above*
Prediction Probability	*>50%*	0	Ave	Ave	Pes	Ave	Pes	Ave	Ave	Pes
	50%	0	Ave	Pes	Ave	Ave	Ave	0	0	Ave
	<50%	Ave	Ave	Ave	Opt	Ave	Opt	Ave	0	Ave

Self-Reported Impact of Contextual Information on Prediction Behavior

Participants self-reported evaluations of the impact of the consequence and likelihood context information revealed similar results as those found for symmetric distributions (Table 13). No evidence was found for differences between the two visualization conditions.

Table 13: Number of participants who reported that they agreed with a statement that the additional information provided by the two context conditions impacted their time predictions for right-skewed distributions

		Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*	***Total***
Consequence	*Median-only*	3	6	4	35	31	79
	Boxplot	3	5	3	30	38	79
Likelihood	*Median-only*	0	0	2	25	52	79
	Boxplot	0	1	1	29	48	79