ABSTRACT

Title of Dissertation: THE QUALITY OF EXPERT JUDGMENT: AN INTERDISCIPLINARY INVESTIGATION
Yashika Forrester, Doctor of Philosophy, 2005

Dissertation Directed by: Dr. Ali Mosleh
Department of Mechanical Engineering-
Reliability Engineering Program

The potential impact of expert judgment on vital tasks, contributes to the desire of decision makers, to know the quality of expert judgment. For years, decision makers and stakeholders have struggled to find answers to "How accurate are experts" and "How close are the experts' estimates to the true values of quantities". Most models or tools used for the prediction of expert performance accuracy and estimates are based on historical performance records of individual experts. Decision makers are often limited to the knowledge of the attributes of the experts and their associated estimates.

This dissertation focuses on two frameworks: (1) to estimate the true value of an unknown quantity, given the estimate of an expert, and (2) to effectively predict expert performance based on the attributes or qualifications of an expert. An extensive meta-analysis of validated expert judgment literature was conducted. The analysis identified the most commonly recommended attributes and evaluated the strength of association between attributes and expert performance.

Results from the analysis demonstrate nonlinear multiple regression relationships between the attributes of experts and their resulting performances. The validation case studies show that the empirical regression equation was effective in forecasting 50% of

the elicited experts' ability to provide accurate responses within 5% of their actual performance. Also, the model predicted 75% if the experts' performance with 15% of actual scores. . The results of the demonstrate that the equations derived to predict performance based on attributes are effective, and can be used to inform decision makers of the expected performance of their experts.

Results also demonstrated that the Bayesian equations developed to predict the true values of unknown quantities based on the estimates of experts are moderately effective. In the validation studies revealed a wide range of possible values for a given quantity, a result influenced by the large variance in the distribution from error.

THE QUALITY OF EXPERT JUDGMENT: AN INTERDISCIPLINARY INVESTIGATION

by

Yashika Forrester

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Committee:
Professor Ali Mosleh, Chair
Professor Carol Smidts
Professor Joseph B. Bernstein
Professor Marvin Roush
Professor Greg Baecher, Dean's Representative

This Dissertation is dedicated in the loving memory of my grandmother, Violet Donaldson-Jackson, an advocate for education, an extraordinary woman, beyond her time.

ACKNOWLEDGEMENTS

I would like to give honor to my LORD and Savior Jesus Christ, who has abundantly blessed me beyond measure. HIS grace is more than sufficient for me.

This manuscript was possible through the efforts and support of a number of individuals. To Professor Ali Mosleh, my mentor, I express my sincere appreciation for his guidance and support in undertaking the research and the writing of this dissertation. I extend special thanks to Dr. Katherine Walker for the permission to use the results of her Benzene Concentration Expert Judgment Case Study in Chapter 6. Also, special thanks to my colleagues at the US Department of Agriculture, especially Dr. Richard Fite and Dr. Eric Grant who have consistently supported and encouraged me throughout this endeavor.

Furthermore, I convey my heartfelt gratitude to my family and friends for their constant prayers and encouraging words, which assisted in making this effort possible. I owe my fondest appreciation to my mother, Ivis Forrester has been patient and supportive. Also, I count myself privileged to have the love and support of my aunt Marcia Vickers, my sister Renee Forrester and wonderful friends, Deon Edward, Claudette Pennant, Claudette Polk, Rita and Mason Wingate as well as Billie and Gregory Malcolm, and Deborah Peace. Again, an earnest thanks to everyone.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Historically, decision-makers utilized inherently subjective expert judgment, such as opinions, predictions, and forecasts (terms frequently used interchangeably) to supplement insufficient data. Relatively cheap and virtually inexhaustible (Cook 1991), they inform critical decision making processes, and significantly impact life and death, as well as financial, spiritual, legal, and social issues. Subjective judgments contain degrees of uncertainty often expressed as error or accuracy. Scientific evaluations of the accuracy of these judgments, on specific issues within disciplines and on broader interdisciplinary issues, yield values across the entire spectrum. These variations depend primarily on the proficiency of the expert or professional, and on the maturity of the technology or the field.

The inherent uncertainty of experts' opinions is the subject of numerous debates. As described by the RAND Corporation, opinions are a hybrid of knowledge and speculation. Knowledge is defined as highly confirmed assertions, and speculation denotes conclusions with little or no supporting evidence (Dalkey 1969). Uncertainty in opinion simply means that, given current knowledge, there are multiple possible states. When appropriate data are not available, the assessment of uncertainty becomes a judgment (Stewart 2000).

Wright and Bolger (1992) have shown that experts have special characteristics that permit improved performance relative to non-experts (Wright and Bolger 1992, Rowe and Wright 2001). When compared to novices, experts make conclusions on relatively few but specific findings, ignoring irrelevant information, while novices examine all details. Empirical studies in many disciplines, including auditing, financial analysis, and product choice, support this premise of novices and experts. Elstein et al. (1978) concluded that experts utilize approximately two thirds of the available data to form judgments, in clinical settings, and spend minimal time seeking confirmation (Elstein et al. 1978, Cox 2002). In addition, cognition studies of expert judgment find information processing less costly for experts than for novices (Camerer and Johnson 1999). However, the challenge still remains how to empirically express and predict expert performance.

The assessment of unique expert performance values across all disciplines is hindered by two primary factors. First, it is unclear in most disciplines what attributes are needed to qualify an individual as an expert. Several authors have proposed their taxonomy for identification and selection, but very few intra- or interdisciplinary standards exist. Frequently based on vague or arbitrary criteria and intuitions, expert selection is often made by a single undefined quality such as extensive knowledge or substantial experience or being renown in the field. These and similar attributes are unfortunately left open to individual interpretation. Within a particular subject area, the lack of a universal standard for expert qualification results in inconsistent identification and selection of experts, and highly flawed aggregated judgments. Second, there are varying published values for expert performance. In fact, across disciplines expert precision ranges from low to high, and 0 to 100 percent. Furthermore, when compared to other tools for assessing evidence, experts performed ‘better than”, “worse than”, and “similar to” mathematical models, chance, and novices, respectively. The root causes of these seemingly random values for expertise stem from several sources. The age of a technology, field, or issue could affect the ability of experts to predict outcomes. Also, the number of years of experience and the depth of knowledge are all factors contributing to variations in expert accuracy.

The voluminous published literature and works on the accuracy of experts indicate a need to more effectively evaluate and compare judgmental accuracy. The aim of this research is three-fold. First, to conduct extensive review and meta-analysis of expert-related literature. Meta-analysis is the process of performing analysis on amalgamated results of various published studies into a common metric. Publications with descriptions of suitable experts, methodologies for selection of experts, and the solicitation and aggregation of their judgments were chosen for review. Accuracies and expert attributes from published case studies were correlated to determine expert performance. Attributes consisted of publications, organizational membership, academic background, practical experience, and peer nomination. The second aim was to develop an empirical mathematical relationship to predict or forecast expert performance or accuracy based on selected expert attributes. Third, develop a Bayesian relationship to predict the true value of an unknown quantity from an expert's estimate, using meta-analysis results. Correlations and other statistical measures were performed using the *SPSS 13.0* software. Empirical equations were validated using two primary and secondary case studies.

CHAPTER 2

LITERATURE REVIEW

2.1 ABSTRACT

Decision makers are often perplexed about expert judgment quality, and degree of confidence to place in experts. This results from a host of issues that affect expert judgment quality, and subsequently leads to controversies. The issues can be broadly classified as ones associated with the individual expert (i.e. attributes, expert definition or distinction), the actual estimates or judgments, as well as, the elicitation process (formal vs informal elicitation), technical aggregation and calibration (aggregation, performance measures of experts and expertise), and historical documentation (limitations of published validation studies).

With much difficulty, scientists have attempted to resolve individual and multiple issues within this quagmire, which contributes to the quality of expert judgment. They have developed several expert calibration models and taxonomies to evaluate expert judgments. However, the effectiveness of these models is limited by the need for historical evidence of an individual's performance. Under most real-life elicitation conditions, decision makers are familiar with the attributes and estimates of experts. A review of the major issues impacting expert judgment quality is presented in this chapter.

2.2 USE OF EXPERT JUDGMENT

The use of expert judgment for scientific purposes is well known. Employed to fill data gaps or to supplement insufficient empirical data, expert judgment serves critical functions to include medical diagnosis, projects duration and estimation costs, as well as engineering design, and risk assessments. The earliest methodologies for formal scientific expert judgment use were developed by the Research and Development (RAND) Corporation, to facilitate decision making, between World War II and the Vietnam War. The first two methods were the Delphi Approach and Scenario Analysis, both developed through the collaborative project of the RAND Corporation, the U.S. Air Force, and Douglas Aircraft, in 1946. The Delphi Method is a group forecasting or consensus technique, whose application spans numerous disciplines. The methodology involves an iterative process of eliciting expert estimates and providing aggregate responses to group, until realistic consensus occurs. Scenario Analysis, on the other hand, is a process of analyzing possible future events by considering alternative possible outcomes or scenarios. Designed to allow improved decision-making, scenario analysis facilitates more complete consideration of outcomes and their implications.

The RAND Corporation has scientifically used expert judgment to assist the U.S. Air Force in identifying and planning for future security. Other organizations, agencies and individuals have followed RAND's ground-breaking lead. Also used for issues relating to human health, scientific expert judgments are employed to evaluate the inhaled radionuclide effect from the nuclear accident on Three Mile Island, and projected future worldwide incidences of acquired immune deficiency syndrome (AIDS). Today, scientific expert judgments are commonly used in part, at the Nuclear

Regulatory Commission, National Aeronautics and Space Administration, the Environmental Protection Agency, other U.S. government agencies, as well as, by medical professionals, weather and sports forecasters.

According to Booker and Meyer (1996), expert judgment is typically used in two fundamental ways. First, it is utilized in the structuring of technical problems. This includes the determination of relevant data for analysis, such as input and output variables, as well as appropriate analysis methodologies and assumptions. Statisticians frequently use their expert judgment in this way. Second, judgments are commonly employed to provide estimates. Experts qualitatively and quantitatively estimate failure or incidence rates, characterize uncertainty, and determine weighting factors for combining data sources (Booker and Meyer 1996).

The primary role players in the use of expert judgment include experts, decision makers, and the public. Experts provide judgments based on their experience and knowledge. The pubic or media rely upon the expertise of the experts and the informed assessment of the decision maker. In addition, the public must sometimes decipher "competing expert claims in the absence of any clear-cut standard to judge these claims" (Munnichs 2004). Therefore, the knowledge of expert judgment quality is essential to all role players.

2.3 DEFINITION OF "EXPERT"

Effective evaluation of expert judgment quality must first consider the multiple definitions of "expert(s)". Experts are partially characterized from generic, scientific, and legal perspectives. They are generally described as individuals who carry out a specified set of tasks expertly (Weiss et al. 2003), or who possess superior knowledge or skill in a particular subject or field. Also, they are referred to as experienced predictors in a domain and have appropriate social or professional credentials (Camerer and Johnson 1997). Cox (2002) and Lesgold et al (1988) defined experts as high-speed recognizers of abnormalities, and diagnostic classifiers who use a personal, organized, perceptual library linked into case-based knowledge. In *Daubert vs. Dow Pharmaceuticals*, the U.S. Supreme Court classifies legal experts in Federal Rule of Evidence 702 as individuals with scientific, technical, skill, experience, training, or education that will assist the trier of fact to understand the evidence or to determine a fact at issue.

Dreyfus and Dreyfus (1986, 1996) summarized the qualitative and quantitative distinctions inherent in differing levels of skillfulness. They concluded that the expert has high levels of procedural knowledge and skills (knowing how) as well as declarative knowledge (knowing what), and contextual flexibility (knowing when and where). In addition, the judgments of experts regarding an action or quantity are independent of principles or rules to facilitate understanding of the situation. The comfort of experts in their domain(s) results in their lack of awareness of the skills being invoked during the process (Dreyfus and Dreyfus 1986, 1996). On the contrary, as declarative and procedural knowledge in the expert are automated to a large extent, there is additional cognitive space available to consider alternatives and reflect upon progress. Thus, an expert seems capable to consider more possibilities and can focus attention on the key information more effectively (Dunphy and Williamson 2004).

2.4 EXPERT ATTRIBUTES

Numerous publications have affirmed the association of specific expert attributes to the quality of expert judgments and performance. Attributes are the characteristics, traits, and peculiarities relating to an individual. Specific attributes are used to distinguish between experts and novices. However, the selection of experts with appropriate attributes or qualifications is subjective.

Most individuals have a unique perception of the quality attributes that make a person an expert. Weiss and Shanteau (2003) identified experts by attributes such as self-proclamation or peer nomination as well as by experience, titles, and degrees. Additional attributes include membership in professional organizations, number of expert participants and characteristics of tasks such as frequency of occurrence and difficulty (Stewart et al. 1997).

Attributes were further classified into tasks-related and perceptual expertise categories. Weiss and Shanteau (2003) suggested four tasks-related categories of expertise. First, those requiring expert judgment for determining awards medals, auditing, grading, and diagnoses. Second, are the predictors such as forecasting the weather, hiring personnel, medical recommendations, advising, and behavioral pattern predictions. Third, the expert instructors who train novices develop computationally aided expert systems, set criteria for testing, or mentor aspiring experts. Fourth, performance experts, who perform beyond the skill of the masses as in playing an instrument, fixing or shooting a basketball, or painting a landscape (Weiss and Shanteau 2003). The perceptual expertise category outlined by Cox (2002) requires attributes of alertness and persistence to ensure clinical accuracy. Each clinician, as a self-aware participant-observer (SAPO) keeping track of what they're thinking "as it happens", are capable of studying their perceptual accuracy, interpretation, pattern matching, judgment and motivation (Cox 2002).

2.5 NUMBER OF EXPERTS NEEDED FOR MAXIMAL ACCURACY

Speculations made about the correlation between expert accuracy and the number of experts used in a study, lead many to conclude the number of experts is directed proportional to accuracy. Questions still remain; is this true? And, with seldom unlimited funding, are there a minimal number of experts needed to obtain optimal accuracy? Hogarth's (1978) normative model suggested that maximal accuracy can be obtained with 6-10 experts, and Ashton's (Ashton & Ashton 1985; Ashton 1986) empirical work as well as many of the studies reviewed by Clemen (1989) showed that between three and six experts lead to high accuracy levels. Some research suggest however, that gains in accuracy are attributed to the inter-correlations of the experts, and minimal gain in accuracy is achieved from redundancy in experts (Johnson et al. 2000, Budescu and Rantilla 2000).

2.6 ELICITATION

Elicitation methods affect judgmental accuracy. Categorized into formal and informal methodologies, expert elicitation methods may range from simple to complex processes. These methods comprise of data gathered implicitly and explicitly, clinically and experienced-based, as well as intuitively, arbitrarily, by guesstimates and gut feelings (Armstrong 1985, pg 73). Formal elicitations are superior to those informally gathered.

A scientific judgment is defined to be informal when there is no formal training provided to debais the expert and the documentation is lacking. Formal expert elicitation (commonly referred to as expert judgment elicitation) refers to a structured procedure designed to gather knowledge about a discipline or area of endeavor from individuals considered human experts in that domain (DeWispelare et al. 1995). Compared to an informal expert judgment process, formal expert elicitation increases the credibility and defensibility of the judgments because of carefully documentation of each expert's rationale and enhances the communication of the results (DeWispelare et al. 1995).

Informally elicited judgments are obtained through unstructured approaches which lack adherence to established protocol or scientific principles. Examples of these types of judgments include intuition, "guesstimations," arbitrary guesses, and gut feelings (Armstrong 1985, pg 73). The most commonly used method for forecasting decisions in conflicts is using their unaided judgment. This is not surprising, as unaided-judgment forecasts can often be derived quickly and cheaply. The simplest and most common of these methodologies entails merely asking for an individual's judgment. Most people have poor intuitions regarding numerical probabilities. Consequently, this inquiry yields the least reliable expert performance results, especially for persons unfamiliar with probability concepts (Cooke 1991). Among theoreticians, the most popular technique is betting rates, introduced by Ramsey (1931), and De Finetti et al (1964). Other informal elicitations include those generally obtained by court testimony from psychologists and psychiatrists about the mental health of defendants (Faust and Ziskin, 1988).

In contrast, formal elicitation is a "well-established probabilistic risk assessment (PRA) tool involving several role-players such as clients, experts, and the project personnel such as data gatherers" (Meyer and Booker 1990, pg 58). It is also defined by Santori et al (2004) as a "heuristic process to obtain evidence and data or information on issues/problems of concern". Formal elicitations can be classified into four major primary categories: indirect, direct, parameter estimation and visual (Jenkins, unpublished). Each method carries various advantages and disadvantages, making them case appropriate. Indirect permits the use of experts not trained in probability concepts and direct is more suitable for those who are familiar to PRA. Parameter Estimation translates data easily into probabilistic results and visual is useful for low information processing demands. Additional and more explicit methods are taxonomic, free-recall, protocol analysis, repertory grid and multidimensional scaling (Jenkins, unpublished).

The earliest methodologies for formal scientific use of expert judgments was developed by the Research and Development (RAND) Corporation, to facilitate decision making, between World War II (WWII) and the Vietnam War. The first two methods were the Delphi Approach and Scenario Analysis. They were both developed through a RAND Corporation joint project with the US Air Force and Douglas Aircraft in 1946. The Delphi Method is a group forecasting technique, generally used for future events such as technological developments, that uses estimates from experts and feedback summaries of these estimates for additional estimates by these experts until reasonable consensus occurs. It has been used in various software cost-estimating activities, including estimation of factors influencing software costs. Scenario Analysis is a process of analyzing possible future events by considering alternative possible outcomes (scenarios). The analysis is designed to allow improved decision-making by allowing more complete consideration of outcomes and their implications.

This process can result in the formation of quantitative estimates for the frequency of physical characteristics of phenomena when the required data is sparse and when the subject is too complex to adequately model.

According to (Booker and Meyer, 1990), there are four primary traits of Expert Judgment. First, “expert judgment is affected by the process of gathering it. Second, expert judgment has uncertainty, which can be characterized and subsequently analyzed. Third, expert judgment can be conditioned on various factors. These factors include: the phrasing of the question, the information the experts considered, the expert’s methods of solving the problem, and the experts’ assumptions. A formal structured approach to elicitation gives analysts a better handle on conditioning effects. Fourth, expert judgment can be combined with other data (via Bayesian approaches).

2.7 QUALITY OF EXPERT ESTIMATES/ JUDGMENT: EPISTEMIC AND ALEATORY EXPERT JUDGMENT

Scientific or informed expert judgment is subjective, and inherently uncertain. Uncertainty in judgments is a function of tasks and attributes of individual experts. Attributes-related uncertainty can be classified as aleatory or epistemic, or combination of both, reflecting the experts limited knowledge on the subject. Aleatory uncertainty is irreducible and results from random or inherent variation. Daneshkhah (2004) believed that aleatory uncertainty arises because of natural, unpredictable variation in the performance of the system under study. Epistemic uncertainty, on the other hand, is reducible and stems from lack of knowledge. Structured elicitation techniques and decomposition of elicitation issues to specific expertise are effective in reducing epistemic uncertainties.

2.8 AGGREGATION AND CALIBRATION OF EXPERT JUDGMENT MEASURES

2.8.1 Performance Measures of Experts and Expertise

The precision of expert judgment is commonly expressed by performance measures. Disciplines differ in their preferred expression of expert judgment accuracy. In the biological and medical fields, the most widely used performance measures are sensitivity, specificity, positive predictive value and negative predictive value. In the engineering and physical sciences, overall accuracy, error, or efficiency is preferred. Calculation of sensitivity, specificity, predictive values, and likelihood ratios requires knowledge about the presence or absence of a specific disorder and require that a test result be dichotomized as either positive or negative. However, when ordering a test in the clinical environment, a clinician does not know whether the disorder is actually present or not, and test results are often ordinal or continuous (Mrus, 2004).

Other natural measures in many fields are predictive (Camerer and Johnson 1997) and differential accuracy, and expert self-confidence rating (Penrod and Cutler 1995). Performance measures terms denote unique functions; however, they are frequently and erroneously used interchangeably.

Scientists have sought to standardize the expression of expertise across all disciplines. “When it is clear that an outcome measure captures expertise, it is appropriate to use it as a means to identify the expert” (Weiss and Shanteau 2003). Cochran, Weiss and Shanteau argued that evaluative skill is the basic cognitive ability that characterizes all these areas of expertise. As a result, they created the CWS index of expertise, ($CWS\_index = \frac{discri\min ation}{inconsistency}$), a ratio of discrimination over inconsistency. “Discrimination refers to the judge’s differential evaluation of the various stimuli similarly over time. Consistency refers to the expert evaluation of the same stimuli similarly over time; inconsistency is its complement” (Weiss and Shanteau 2003).

Experts’ confidence in their judgments is widely used by decision-makers to assess accuracy. However, this must be viewed with much caution. Researchers agree the correlation “between identification accuracy and confidence in identification judgments is weak.” Penrod and Cultler (1995) conducted a review of literature “that there is a weak association between eyewitness confidence and identification accuracy.” The results this article reinforced the conclusion and caution by many researchers that the correlation is weak.

Another technique for measuring expert accuracy is the receiver operating characteristic (ROC). In this approach, an expert makes a judgment about the occurrence of an event, viewed from two perspectives: the expert as a signal detection system; and the individual utilizing the expert opinion. The ROC fully describes the potential ability of the expert to distinguish between the possible occurrences of an event. The models assert that the expert compares the strength of the evidence to one or more decision criteria. The values of these decision criteria are influenced by two factors: The expert's belief about the prior probability of the event and goals the expert has in making a decision (Harvey, 1992).

2.8.2 Aggregation of expert judgments or opinions

Methods of expert opinion aggregation vary from simple geometric (Xu 1999) and, arithmetic aggregations, to weighted tools, to more complex methods that account for dependency. Mosleh (1992) suggested a Bayesian approach to combining expert opinions when there is an inherent variability among the estimates. Several authors (Cooke 1990; Genest and Zidek 1986; French 1985) provided literature reviews on Bayesian models for aggregating expert opinion (Mosleh 1992). An iterative procedure for approximating the optimal consensus of expert opinions was introduced by Hsuan-Shih Lee (2001).

In another study, the opinions of experts and the public were aggregated using the analytic hierarchy process (AHP) and multi-attribute utility analysis (MAUA), and for uncertainty analysis, a fuzzy set based approach was adopted (Sohn 2001). Moon and Kang proposed a technique that utilizes and fuzzy set theory in the aggregation of expert judgments. In the techniques, two main key concepts are employed: linguistic variables and fuzzy numbers. Linguistic variables first represent the relative importance of evaluation criteria under consideration and the degrees of confidence on each expert perceived by the decision maker, and then are replaced by suitable triangular fuzzy numbers for arithmetic manipulation (Moon 1999).

2.8.3 Expert Calibration/ Distinction

This study finds that there is little empirical evidence for the propositions (1) that experts judge risk differently from members of the public or (2) that experts are more veridical in their risk assessments. Methodological weaknesses in the early research are documented, and it is shown that the results of more recent studies are confounded by social and demographic factors that have been found to correlate with judgments of risk. Using a task-analysis taxonomy, a template is provided for the documentation of future studies of expert-lay differences/similarities that will facilitate analytic comparison (Rowe and Wright 2001).

The goal is to derive an empirical measure of expert judgment. The tasks that experts do were partitioned into four categories, with evaluation being the primary function underlying all expertise. Viewing the evaluator as a measuring instrument, we propose that two necessary characteristics of an expert are the ability to discriminate among different stimuli in the domain and to be consistent in judgments of the same stimulus. We combined measures of those characteristics to form a ratio, the CWS index of expertise (Weiss and Shanteau, unpublished).

Firstly, the many characteristics of expertise were examined: they included aspects of pattern recognition, knowledge, skill, flexibility, meta-cognitive monitoring, available cognitive space and teaching abilities. Secondly, three educational models from different models from different domains (nursing, surgical education, education) are analyzed, compared and contrasted, in relation to both educational approach and the development of expertise. Thirdly, a new model for the development of expertise is proposed, incorporating aspects of each of the three previously discussed models.

Within this new model, four phases of development are proposed, culminating in the achievement of expertise. Furthermore, it is noted that under certain circumstances performance can deteriorate, and that with appropriate support, there can be recursion back through earlier phases of development (Dunphy and Williamson, 2004).

2.9 PUBLICATIONS OF VALIDATED EXPERT JUDGMENT STUDIES

2.9.1 Studies with accuracy information only

Brancato et al (2002) reviewed the accuracy of expert judgment in predicting breast cancers. In this study, cases wherein invasive surgery was performed after expert radiologists made judgments of mammograms and ultrasonographs were compared with mathematical models and definitive results from surgery. The results showed that neither approach reached a satisfactory accuracy, but the radiologist's judgment (sensitivity 97.1%, specificity 81.9%, positive predictive value 98.4%, negative predictive value 71.6%, overall accuracy 95.8%) was slightly superior to the mathematical model (sensitivity 93.2%, specificity 87.9%, positive predictive value 98.8%, negative predictive value 53.7%, overall accuracy 92.8%) (Brancato et al. 2002).

2.9.2 Studies with published quality attributes and accuracy information

A study by Margo (2000) compared the inter-rater and inter-group agreement in judging physician maloccurrence and compliance with standards of care using the implicit case review process. The survey was mailed to 140 board-certified opthalmologists and 140-board certified ophthalmologist with fellowship training, and the process designed wherein the identity of the physicians could not be determined. In the mailed packages, the complete medical records of two patients were given and the participants were asked after reviewing the information to answer the following three questions:

(1) Was the clinical outcome in the case attributable to the medical care the patient received (an error of either commission or omission) for ophthalmologist labeled “A”?

(2) Did ophthalmologist “A” meet the standard care expected of him or her in this clinical situation?

(3) If the standard of care was not met, was the deviation minor, moderate, or major?

(The answers were coded so they could not be traced to individual participants) There was good within-group agreement for finding clinical error in management and not meeting the standard error of care. Overall, only 35% of the respondents believed that the ophthalmologists in the case reviewed committed an error of either commission or omission, while 45% of the reviewers believed that the physicians did not met the standard of care. The fact is that the ophthalmologists did commit an error of commission or omission; on the other hand, they did meet the standard of care (Margo 2000).

Haber and Haber (2003) reviewed the three kinds of available data about the accuracy of fingerprint comparisons made by human latent fingerprint examiners: the accuracy of consensus fingerprint comparisons made by groups of examiners working in crime laboratories; the proficiency and certification test scores of latent fingerprint examiners tested individually; and the results of controlled experiments on the accuracy of fingerprint comparisons. Their study showed that consensus judgments of fingerprint comparisons show either in determinant or large error rates of 14% for examiners acting alone and 2% for examiners in groups. Furthermore, the proficiency and certification procedures in current use lack validity, and cannot serve to specify the accuracy or skill level of individual fingerprint examiners (Haber and Haber, 2003).

2.10 SUMMARY

The utilization of expert judgments for critical functions necessitates the fundamental need for expert precision. Researches have found it difficult to locate and aggregate studies with values of expert accuracy/performance. Expert accuracy values reveal an array of values, from low to high extremes. These variations are the result of many factors. Some of the most significant factors include the type of elicitation, number of experts, the methodology for judgment aggregation, and expert attributes. Some studies publish detailed descriptions of experts and some publish vague accounts. Accuracy values cannot be effectively aggregated because there are varying criteria or requirements for experts in a particular field. In addition, there are limited approaches/ methodologies to measuring expert accuracy across all disciplines.

The varied criteria for expert identification and selection make aggregation of responses difficult. Many characteristics of the expert are not detailed, which make it difficult to aggregate data within and across disciplines as well as to determine the validity or accuracy of the estimate.

CHAPTER 3

CORRELATION OF EXPERT ATTRIBUTES TO PERFORMANCE: AN INTERDISCIPLINARY INVESTIGATION

3.1 ABSTRACT

The use of expert judgment to supplement insufficient data for critical decisions prompts the need to know expert accuracy. The literature reports several validated expert judgment studies across disciplines. These studies focus on narrow scopes such as expert accuracy in the diagnosis of specific diseases, or the prediction of particular event outcomes. Several factors including the definition and attributes of experts impact accuracy. The definition of an expert is subjective, as evidenced by the multiple existing definitions. As a result, the identification and selection criterion for experts in most disciplines are inadequate, leading to ambiguous qualification attributes and unpredictable performance outcomes.

This chapter focuses on the meta-analysis of published interdisciplinary expert judgment accuracy and error case studies. The meta-analysis treated all the interdisciplinary case studies as one unit. The results of the meta-analysis show beta and right truncated normal probability density function fits for the overall percent accuracy of experts, and a logarithmic and gamma probability density function fits for the mean absolute percent error of experts. In addition, the correlation coefficients of individual attributes to accuracy were determined.

3.2 INTRODUCTION

The utilization of expert judgments for critical decisions and functions has prompted the fundamental need to know the precision of experts. In the absence of empirical data, expert judgment is used to aid decision making, and inherently contains degrees of uncertainty. Experts are routinely selected based on the decision makers' predetermined perception of appropriate characteristics or attributes, along with their confidence in the experts' ability to accurately provide judgments. The resulting judgments of these experts are widely and sometimes indiscriminately used without knowledge of their quality or accuracy (Kane 1995, and Roebber et al. 1997).

Empirical evidence of expert performance is often desired and seldom attained. Most tools designed to aid decision makers to effectively evaluate an individual expert's judgment are developed from the historical performance data of that expert. In most realistic elicitation situations, the only available information is the attributes and estimates of experts. Therefore, an empirical framework wherein expert performance is derived essentially from attributes and estimates is necessary.

There exists an abundant assortment of models to evaluate expert performance within the Bayesian framework. The justification(s) for models utilized, and the allocation of empirical values to the respective model parameters are theoretical. In practice, decision makers remain confounded about the appropriate model and value(s) of associated parameter(s) to represent expert judgment across disciplines. For that reason, a meta-analysis of literature containing validated expert judgment case studies was performed to determine the most practical and appropriate models: (1) to represent the accuracies of experts across all disciplines, and (2) to depict the error in expert estimates, and subsequent model parameters.

The meta-analysis treated the interdisciplinary case studies as two units, one for categorical quantities and the other for continuous. The articles were evaluated for commonly recommended expert attributes, accuracy and error. The percent accuracies and mean absolute percent errors of experts were fitted to probability distributions, and later used in chapters 4 and 5 to develop likelihood functions. Furthermore, the correlation coefficients of each attribute to accuracy were derived, and also used to develop the likelihood functions in chapters 4 and 5. The results of the meta-analysis are presented in this chapter.

3.3 BACKGROUND

Expert judgments, usually reported as opinions, predictions, or forecasts, are terminologies frequently used interchangeably. Utilized in part for life and death, as well as financial, spiritual, legal, and social decisions, they are relatively cheap and virtually inexhaustible (Cook 1991). Expert judgments inform, and significantly impact critical decision making processes.

The precision of expert judgment is commonly expressed by performance measures. Disciplines differ in their preferred expression of expert judgment accuracy. In the biological and medical fields, the most widely used performance measures are sensitivity, specificity, positive predictive and negative predictive values. In the engineering and physical sciences the performance measures of overall accuracy, error, and efficiency, as well as reliability is preferred. Other fields use measures including predictive (Camerer and Johnson 1991) and differential accuracy, and expert self-confidence rating (Penrod and Cutler 1995). Each performance measure term denotes a unique function and is frequently and erroneously used interchangeably. To effectively communicate within a discipline, and between disciplines, a standard measure or taxonomy of measures is needed.

Cochran, Weiss and Shanteau developed the CWS index of expertise to empirically evaluate expert performance. This index is formed on the premise that evaluative skill is the basic cognitive ability that characterizes all areas of expertise. The CWS index of expertise, $CWS\,index = \frac{discri\min ation}{inconsistency}$, is a ratio of discrimination over inconsistency. Discrimination, according to Weiss and Shanteau (2003), refers to the judge's differential evaluation of the various stimuli similarly over time. "Consistency refers to the expert evaluation of the same stimuli similarly over time; inconsistency is its complement" (Weiss and Shanteau 2003). This index is particularly suitable for cases wherein performance records exist.

Rowe and Wright (2001) conducted a meta-analysis of nine empirical studies using a new task-analysis taxonomy to investigate expert-lay judgments of risk. The findings show methodological weaknesses and confounding demographic and social variables in these studies. The taxonomy developed, provides additional information to that previously reported by Slovic (1999), on socio-demographics, the nature of the day-to-day activities of the experts, nature of the risk assessments made by the experts and on the "potential learnability of high quality judgmental performance". Rowe and Wright (2001) contended that the above cited information was not available in the studies investigated; therefore no conclusions could be made about the differences in the quality of expert and lay judgments.

In this chapter, however, the measures of expertise are reduced to either mean absolute percentage error or percent accuracy. The mean absolute percentage error is used to express expertise for continuous quantities. Absolute percentage error is the ratio of the absolute difference of estimated and actual values, over the actual value of the quantity. This value denotes the deviation of the expert from the true value, and was either implicitly stated or extracted from data in each published case study. In contrast, the percent accuracy measure reflects expertise for categorical quantities. For example, the expertise of a medical specialist may be reflected in the percentage of accurate diagnosis made.

The literature reports many studies of expert judgmental accuracy evaluations across several disciplines. Individual studies are based on narrowly focused scopes, typically on a particular issue. In the area of cardiovascular diseases for example, Lipinski et al. (2002) found that expert cardiologists were better able to estimate the presence of clinically significant and severe coronary artery diseases than randomly selected cardiologists and internists, 76% and 73%, respectively. The study of Reischman and Yarandi (2002) show that expert nurses compared to novices were significantly better judges in cues utilization for critical cardiovascular care, 72% and 23%, respectively.

Expert performance must consider multiple definitions or qualification criteria for an "expert". Shanteau (1993) fittingly stated that there are almost as many definitions of *expert* as there are scholars in the field. This results, in part, from the lack of standardization in the definition of and qualification for experts, within and across disciplines. Experts are characterized partially from generic, scientific, and legal perspectives. They are generally described as individuals who carry out a specified set of tasks expertly (Weiss et al. 2003), or who possess superior knowledge or skill in a particular subject or field. Also, they are experienced predictors in a domain and have appropriate social or professional credentials (Camerer and Johnson 1999). Cox (2002) and Lesgold et al (1988) defined experts as high-speed recognizers of abnormalities, and diagnostic classifiers who use a personal, organized, perceptual library linked into case-based knowledge. In *Daubert vs. Dow Pharmaceuticals*, the U.S. Supreme Court classifies legal experts in Federal Rule of Evidence 702 as individuals with scientific, technical, skill, experience, training, or education that will assist the trier of fact to understand the evidence or to determine a fact at issue (Penrod et al. 1995).

It is unclear in most disciplines what attributes are needed to qualify an individual as an expert. Attributes, as used in this context, are the characteristics, traits, experiences, and peculiarities relating to an individual. Most individuals have a unique perception of the quality attributes that makes a person an expert. Weiss and Shanteau (2003) identified experts by attributes such as self-proclamation or peer nomination as well as by experience, titles, and degrees. Additional attributes listed by Stewart et al. (1997) include membership in professional organizations, number of expert participants and characteristics of tasks such as frequency of occurrence and difficulty. The perceptual expertise category outlined by Cox (2002) requires attributes of alertness and persistence to ensure clinical accuracy. An established standard of necessary attributes for expert qualification helps to facilitate consistency in the identification and selection of experts.

The validation of expert performance is necessary in the promotion of expert judgment as a useful scientific tool. Experts are commonly validated by known event outcomes (Bariciak et al. 2003, Hughes et al. 1992), empirical research findings (Bentley et al. 2002) and statistical models. A review of literature yielded several validation deficiencies. First, expert judgments are needed in domains where correct answers are frequently nonexistent (Weiss and Shanteau, unpublished). As a result, the experts with inherently uncertain judgments are used to validate other experts. Second, the methodology for validation is vague or not stated. Lastly, expert judgment is heavily used in engineering and physical sciences (Fleming 1991, Mumpower and Stewart 1996), yet there is a paucity of information on validated expert judgment case studies. Also, in other areas such as law, medicine, and forecasting there are several published case studies on validated expert performance.

Knowledge of judgment accuracy is relevant and significant in decision making processes, and to the mainstream acceptance of expert judgment. Limitations to broader focused empirically validated expert performance can be reduced with standardized attributes requirements for experts. In light of the current state of this field, the paper proposes an "attribute-accuracy" taxonomy containing based on correlation strengths.

3.4 METHODOLOGY

The voluminous published literature and works on expert judgments indicate a need to more effectively evaluate the attributes and judgmental accuracy of experts. In this paper, an extensive review and meta-analysis of expert-related literature was conducted. Meta-analysis is a statistical process of combining results of various studies into a common metric and performing analysis. In addition, a preliminary attribute-accuracy taxonomy from ongoing research is presented.

The taxonomy is being developed from publications containing expert descriptions and performance measures of experts. The aim of this work is to identify attributes that are good predictors of expert performance (Dunphy and Williamson 2004). Attributes selected in this study were publications, organizational membership, academic background, and frequency of issues/events, as well as, practical experience, number of experts in study, and peer nomination.

The search for the accuracy of *expert opinion* or *expert judgment* began with a general survey of past and most recent literature, books, internet publications, refereed and non-referred sources. The wide literature search included in part the following databases: WorldCat, Agricola, DOE's Information Bridge, Civil Engineering (CE) Database, Energy Citations Database, Waste Management Research Abstracts, PubMed, and Medline. The most insightful abstracts were found in PubMed, Medline, and WorldCat. In addition, a worldwide exploration of the Dissertation Abstracts database was performed to identify any similar or exact work across all disciplines.

Over 1700 sources, primarily periodical abstracts and books were initially flagged for general relevance. Each source was examined for significance to the elicitation, aggregation, validation and quality of expert judgment or opinion. Of these sources approximately 790 were selected and filed in the following categories: medical, animal, plant, engineering, legislative and forecasting related expert judgment case studies, and theoretical information. From this stockpile of resources, 147 expert judgment case studies were identified and 80 used for the analysis in this paper. The 80 selected case studies were thoroughly examined for each expert attribute listed earlier, along with the expert performance measures. The remaining 67 case studies were temporarily discarded for a variety of reasons. Many discarded studies contained no expert information, only performance scores; others cases contained qualitative performance measures or elusive methodologies for expert validation.

The evaluation of the data included preliminary statistical analysis in SPSS, and subsequent assessment in Bayesian framework. Using the *SPSS 13.0* software, coded entries were made into a spreadsheet and analyzed for each case study. Attributes were coded as continuous, nominal, and ordinal variables, and were subjected to several analyses to determined descriptive statistics, box and scatter plots, histograms, distribution fittings and correlations. Following the appropriate fitting and selection of distributions in SPSS, the parameters of the distributions for error estimated in the Bayesian framework. Selected analyses and results are presented in the next section of this paper.

3.5 RESULTS & DISSCUSSION

3.5.1 Mean Absolute Percentage Error of Continuous Quantities

All validated expert judgment case studies employed in the meta-analysis, were evaluated for two common metrics of expertise. The first metric, absolute percentage error, APE or E_i, was either implicitly stated or extracted from data in each case study. Absolute percentage error is the ratio of the absolute difference of estimated and actual value over actual value.

$$APE = E_i = \left|\frac{u_i^{'} - u}{u}\right| \cdot 100 \qquad (3.5.1.1)$$

The meta-analysis of continuous quantities yielded an array of empirical values for absolute percentage error. Fifty eight data points characterizing error were obtained from literature, and binned in the histogram in Figures 1 and 2, which depict the relative frequency of APE. Furthermore, the histogram was evaluated for characteristics of probability density functions resulting in two close fits, the exponential and lognormal. Figure 1 displays the exponential distribution fit to the histogram representing APE values. The resulting probability density function of the exponential distribution is mathematically represented below:

$$f(E) = \frac{1}{\beta} e^{-\frac{1}{\beta}(E)} \qquad (3.5.1.2)$$

A value of 11.059 was generated by the BestFit Software in Figure 1, for the parameter β. The value for β was also evaluated in the Bayesian framework, from the maximum likelihood function for expert errors given β:

$$L\left(E_1 \ldots E_N \middle| \beta\right) = \prod_{i=1}^{N} \frac{1}{\beta} e^{-\frac{1}{\beta}(E_i)} = \left(\frac{1}{\beta}\right)^N e^{-\frac{1}{\beta}\left(\sum_{i=1}^{N} E_i\right)} \tag{3.5.1.3}$$

The resulting likelihood function in Equation 3.5.1.3 was maximized for β:

$$\frac{d}{d\beta}\left[L\left(E_1 \ldots E_N \middle| \beta\right)\right] = 0 \tag{3.5.1.4}$$

and the MLE for β yields:

$$\hat{\beta} = \frac{\sum_{i=1}^{N} E_i}{N} \tag{3.5.1.5}$$

Below, $\hat{\beta}$ is determined from the substitution of Table 1 data into Equation 3.5.1.5:

$$\hat{\beta} = \frac{\sum_{i=1}^{N} E_i}{N} = \frac{\sum_{i=1}^{58} E_i}{58} = \frac{641.4}{58} = 11.059 \tag{3.5.1.6}$$

<u>Development of Posterior Distribution for β</u>

The MLE obtained in Equation 3.5.1.5, provides a single point estimate for β, the parameter of the exponential function for error. There is however, uncertainty surrounding β. This uncertainty is represented in the following posterior distribution for β, given the expert error evidence:

$$\pi\left(\beta \middle| E_1 \ldots . E_N\right) = \frac{L\left(E_1 \ldots . E_N \middle| \beta\right)\pi_0(\beta)}{\int L\left(E_1 \ldots . E_N \middle| \beta\right)\pi_0(\beta)d\beta} \tag{3.5.1.7}$$

From the likelihood function of error evidence given β in Equation 3.5.1.3:

Let $\lambda = {}^{1}\!/_{\beta}$

$$\pi\left(E_1 \ldots E_N \middle| \lambda\right) = \prod_{i=1}^{N} \lambda e^{-\lambda|E_i|} = (\lambda)^N e^{-\lambda\left(\sum_{i=1}^{N}|E_i|\right)} \tag{3.5.1.8}$$

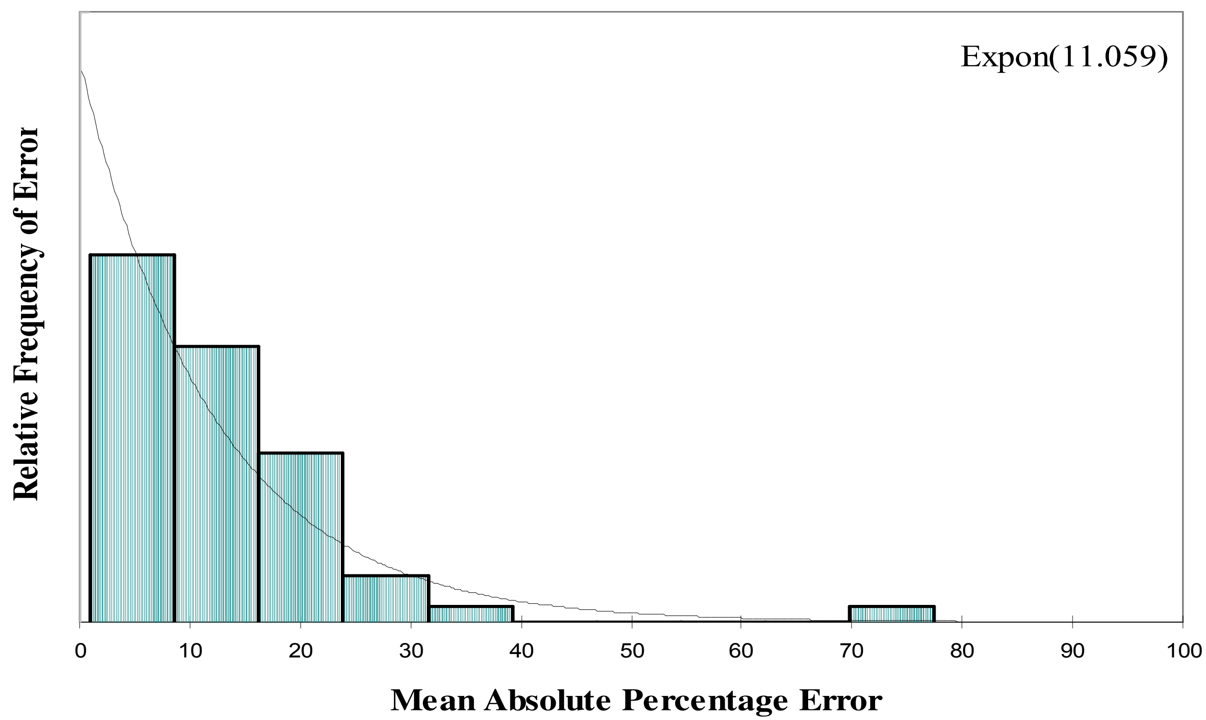

Figure 1. Exponential Probability Density Function Fitted to Relative Frequency of Mean Absolute Percentage Error Values Obtained From Meta-Analysis

Table 1. Meta-Analysis Results of Interdisciplinary Expert Errors

EXPERT ERRORS							
1	26.7	**16**	14.7	**31**	0.9	**45**	8.6
2	28.3	**17**	0.6	**32**	0.1	**46**	11.3
3	14.7	**18**	0.7	**33**	0.1	**47**	16.9
4	16	**19**	1.2	**34**	0.4	**48**	9.1
5	11.7	**20**	1.3	**35**	4.1	**49**	10.2
6	13.8	**21**	1.4	**36**	1.4	**50**	8
7	76.4	**22**	1.5	**37**	0.1	**51**	9.5
8	20.1	**23**	4.6	**38**	0.1	**52**	11.5
9	12.2	**24**	6	**39**	0	**53**	22.7
10	34.1	**25**	7.2	**40**	13.8	**54**	7.6
11	20.3	**26**	8.5	**41**	16.1	**55**	12.8
12	18.6	**27**	0.5	**42**	16.7	**56**	20.8
13	18.1	**28**	0.7	**43**	10.3	**57**	8
14	24.1	**29**	1	**44**	10.1	**58**	7.5
15	16.4	**30**	1.3				

If we assume a gamma prior on λ, then:

$$\pi_0(\lambda) = \frac{b^a (\lambda)^{a-1} e^{-b\lambda}}{\Gamma(a)} \tag{3.5.1.9}$$

Following the substitution of Equations 3.5.1.8 and 3.5.1.9 into Equation 3.5.1.7, the posterior distribution becomes:

$$\pi(\lambda | E_1 \ldots E_N) = k^{-1}\left(\frac{b^a (\lambda)^{a-1} e^{-b\lambda}}{\Gamma(a)}\right)(\lambda)^N e^{-\lambda\left(\sum_{i=1}^{N}|E_i|\right)} \tag{3.5.1.10}$$

or

$$\pi(\lambda | E_1 \ldots E_N) = \frac{(\lambda)^{N-1+a} \, exp\left[-\lambda\left(b + \sum_{i=1}^{N}|E_i|\right)\right]}{\int_0^\infty (\lambda)^{N-1+a} \, exp\left[-\lambda\left(b + \sum_{i=1}^{N}|E_i|\right)\right] d\lambda} \tag{3.5.1.11}$$

and the normalization factor in Equation 3.5.1.11 yields:

$$\int_0^\infty (\lambda)^{N-1+a} \, exp\left[-\lambda\left(b + \sum_{i=1}^{N}|E_i|\right)\right] d\lambda = \frac{\Gamma(N+a)}{\left(b + \sum_{i=1}^{N}|E_i|\right)^{N+a}} \tag{3.5.1.12}$$

The resulting posterior distribution for λ is:

$$\pi(\lambda | E_1 \ldots E_N) = \frac{\left(b + \sum_{i=1}^{N}|E_i|\right)^{N+a} (\lambda)^{N-1+a} \, exp\left[-\lambda\left(b + \sum_{i=1}^{N}|E_i|\right)\right]}{\Gamma(N+a)} \tag{3.5.1.13}$$

or

$$\pi(\beta | E_1 \ldots E_N) = \frac{\left(b + \sum_{i=1}^{N}|E_i|\right)^{N+a} \left(\frac{1}{\beta}\right)^{N-1+a} \, exp\left[-\frac{1}{\beta}\left(b + \sum_{i=1}^{N}|E_i|\right)\right]}{\Gamma(N+a)} \tag{3.5.1.14}$$

Maximizing the posterior distribution of Equation 4.5.2.14 for β:

$$\frac{d}{d\beta}\left[\pi\left(\beta \middle| E_1 \ldots E_N\right)\right] = 0 \tag{3.5.1.15}$$

$$\frac{\left(b + \sum_{i=1}^{N}\left|E_i\right|\right)^{N+a}}{\Gamma\left(N + a\right)} \frac{d}{d\beta}\left[\left(\frac{1}{\beta}\right)^{N-1+a} exp\left[-\frac{1}{\beta}\left(b + \sum_{i=1}^{N}\left|E_i\right|\right)\right]\right] = 0$$

The maximum likely value for β yields:

$$\hat{\beta} = \frac{b + \sum_{i=1}^{N}\left|E_i\right|}{N - 1 + a} \tag{3.5.1.16}$$

The exponential fit discards the lack of error values at and near zero. The distribution assumes experts are most likely to achieve zero error, and larger errors with decreasing probability. On the other hand, in Figure 2, the logarithmic fit

$$f(E) = \frac{1}{\sqrt{2\pi}\sigma_E\left(E\right)} e^{-\frac{1}{2}\left(\frac{ln\left(E\right) - ln\,\mu_E}{\sigma_E}\right)^2} \tag{3.5.1.17}$$

assumes a low probability of error at and near zero. The second distribution appropriately fitted the relative frequency of errors in the histograms in Figures 2 and 3 to the lognormal distribution. The BestFit Software generated a distribution mean and standard deviation of 14.9% and 20.6%, respectively.

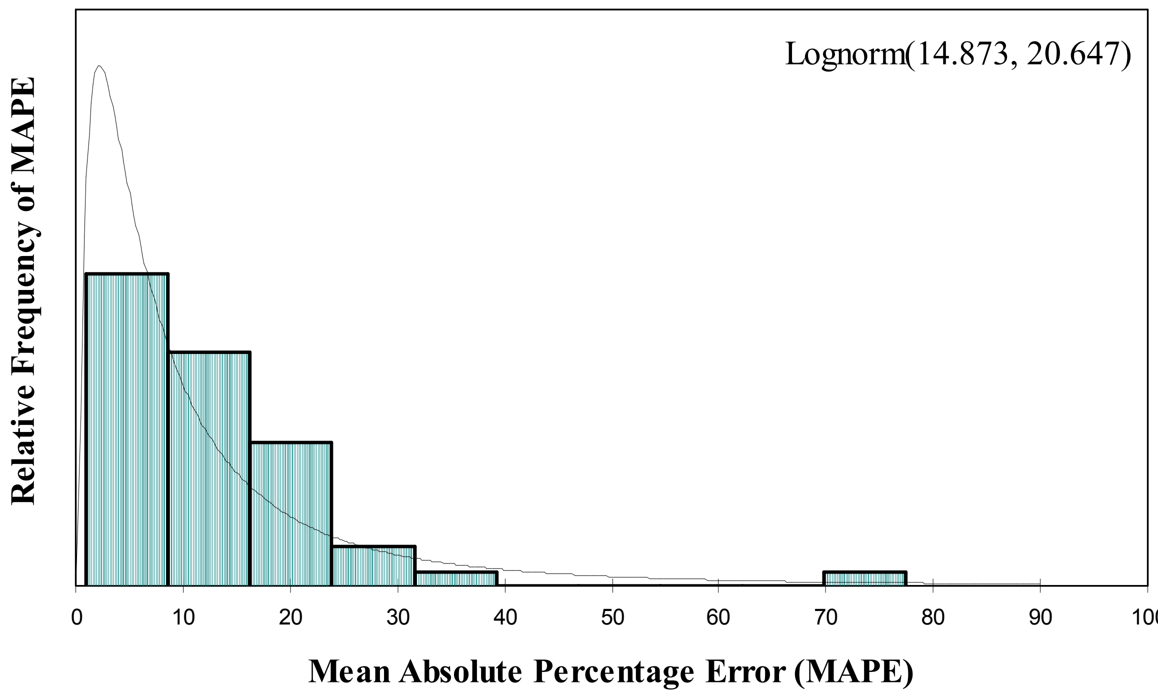

Figure 2. Lognormal Probability Density Function Fitted to Relative Frequency of Mean Absolute Percentage Error Values Obtained From Meta-Analysis

In the ensuing steps, distribution parameters σ and μ, are estimated in the Bayesian framework from the maximization of the likelihood function. The likelihood function depicts the distribution of expert errors, given σ and μ:

$$L(E_1 \ldots E_N | \sigma, \mu) = \prod_{i=1}^{58} \frac{1}{\sqrt{2\pi}\sigma_E(E_i)} e^{-\frac{1}{2}\left(\frac{ln(E_i) - ln\,\mu_E}{\sigma_E}\right)^2}$$

$$L(E_1 \ldots E_N | \sigma, \mu) = \left(\frac{1}{\sqrt{2\pi}\sigma_E}\right)^{58} e^{-\frac{1}{2(\sigma_E)^2}\sum_{i=1}^{58}(ln(E_i) - ln\,\mu_E)^2} \prod_{i=1}^{58} \frac{1}{E_i} \tag{3.5.1.18}$$

Maximizing the likelihood function for μ_E

$$\frac{\partial}{\partial \mu_E}\left[L(E_1 \ldots E_N | \sigma_E, \mu_E)\right] = 0 \tag{3.5.1.19}$$

$$\frac{1}{(\sigma_E)^2}\frac{1}{\mu_E}\sum_{i=1}^{58}(ln(E_i) - ln\,\mu_E) = 0$$

Evaluating Equation 3.5.1.19 for μ_E yields the MLE, $\hat{\mu}_E$:

$$ln\,\hat{\mu}_E = \frac{\sum_{i=1}^{58} ln(E_i)}{58}$$

$$\hat{\mu}_E = exp\left(\frac{\sum_{i=1}^{58} ln(E_i)}{58}\right) = 5.104 \tag{3.5.1.20}$$

Similarly, maximizing the likelihood function for σ_E

$$\frac{\partial}{\partial \sigma_E}\left[L(E_1 \ldots E_N | \sigma_E, \mu_E)\right] = 0 \tag{3.5.1.21}$$

and solving Equation 3.5.1.20 for σ_E yields the MLE $\hat{\sigma}_E$:

$$\hat{\sigma}_E = \sqrt{\frac{\sum_{i=1}^{58}(ln(E_i) - ln\,\hat{\mu}_E)^2}{58}} = \pm 1.64 \tag{3.5.1.22}$$

The likelihood empirical values derived in the Bayesian framework for $\hat{\mu}_E$ and $\hat{\sigma}_E$ will be used to further develop Equations in Chapter 4.

Development of Posterior Distribution for μ_E and σ_E

Although, the maximum likelihood estimators in Equations 3.5.1.19 and 3.5.1.21 supply single point estimate for μ_E and σ_E, for the parameters of the lognormal function, they do not account for associated uncertainties. The uncertainty surrounding β is not reflected in the MLE. Therefore, the following posterior distribution for μ_E and σ_E, given the evidence expert errors was developed:

$$\pi(\sigma_E, \mu_E | E_1 \ldots . E_N) = \frac{L(E_1 \ldots . E_N | \sigma_E, \mu_E) \pi_0(\sigma_E, \mu_E)}{\iint\limits_{\sigma_E, \mu_E} L(E_1 \ldots . E_N | \sigma_E, \mu_E) \pi_0(\sigma_E, \mu_E) d\sigma_E d\mu_E} \qquad (3.5.1.22)$$

Assuming an exponential prior distribution

$$\pi_0(\sigma_E, \mu_E) = \frac{1}{\mu_E} e^{-\sigma_E} \qquad (3.5.1.23)$$

The posterior distribution for a single expert *"i"* is denoted:

$$\pi(\sigma_E, \mu_E | E_i) = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_E(E_i)} e^{-\frac{1}{2}\left(\frac{\ln(E_i) - \ln \mu_E}{\sigma_E}\right)^2} \right) \left(\frac{1}{\mu_E} e^{-\sigma_E} \right)}{\iint\limits_{\sigma, \mu} \frac{1}{\sqrt{2\pi}\sigma_E(E_i)} e^{-\frac{1}{2}\left(\frac{\ln(E_i) - \ln \mu_E}{\sigma_E}\right)^2} \left(\frac{1}{\mu_E} e^{-\sigma_E} \right) d\sigma d\mu} \qquad (3.5.1.24)$$

which reduces to

$$\pi(\sigma_E, \mu_E | E_i) = \frac{2e^{-\sigma_E}}{\sqrt{2\pi}\mu_E \sigma_E} e^{-\frac{1}{2}\left(\frac{\ln(E_i) - \ln \mu_E}{\sigma_E}\right)^2} \qquad (3.5.1.25)$$

Likewise, the posterior distribution for <u>multiple experts</u> is denoted:

$$\pi\left(\sigma_E,\mu_E \middle| E_1 \ldots E_N\right)=k^{-1}\left(\frac{1}{\mu_E}e^{-\sigma_E}\right)\left(\frac{1}{\sqrt{2\pi}\sigma_E}\right)^N \bullet \\ e^{-\frac{1}{2(\sigma_E)^2}\sum_{i=1}^{N}(ln(E_i)-ln\,\mu_E)^2}\prod_{i=1}^{N}\frac{1}{E_i} \tag{3.5.1.26}$$

or

$$\pi\left(\sigma_E,\mu_E \middle| E_1 \ldots E_N\right)= \\ \frac{\left(\frac{1}{\mu_E}e^{-\sigma_E}\right)\left(\frac{1}{\sigma_E}\right)^N e^{-\frac{1}{2(\sigma_E)^2}\sum_{i=1}^{N}(ln(E_i)-ln\,\mu_E)^2}}{\iint_{\sigma_E,\mu_E}\left(\frac{1}{\mu_E}e^{-\sigma_E}\right)(\sigma_E)^{-N}e^{-\frac{1}{2(\sigma_E)^2}\sum_{i=1}^{N}(ln(E_i)-ln\,\mu_E)^2}\,d\mu_E d\sigma_E} \tag{3.5.1.27}$$

Integrating Equation 3.5.1.27 over μ_E

$$\pi\left(\sigma_E,\mu_E \middle| E_1 \ldots E_N\right)=\frac{\left(\frac{1}{\mu_E}e^{-\sigma_E}\right)\left(\frac{1}{\sigma_E}\right)^N e^{-\frac{1}{2(\sigma_E)^2}\sum_{i=1}^{N}(ln(E_i)-ln\,\mu_E)^2}}{-\int_{\sigma_E}\left(e^{-\sigma_E}\right)(\sigma_E)^{-N+1}\left(\frac{N}{2}\sqrt{2\pi}\right)d\sigma_E} \tag{3.5.1.28}$$

The resulting posterior distribution in Equation 3.5.1.28 is subsequently integrated over σ_E, yielding:

$$\pi\left(\sigma_E,\mu_E \middle| E_1 \ldots E_N\right)=\frac{2e^{-\sigma_E}\,exp\left[-\frac{1}{2(\sigma_E)^2}\sum_{i=1}^{N}(ln(E_i)-ln\,\mu_E)^2\right]}{-\left(N\mu_E\sqrt{2\pi}\right)(\sigma_E)^N\,\Gamma(-N+2)} \tag{3.5.1.29}$$

Maximizing the posterior for σ_E

$$\frac{\partial}{\partial\sigma_E}\left[\pi\left(\sigma_E,\mu_E\middle|E_1\ldots E_N\right)\right]=$$

$$\frac{\partial}{\partial\sigma_E}\left[\frac{2e^{-\sigma_E}\,exp\left[-\frac{1}{2(\sigma_E)^2}\sum_{i=1}^{N}\left(ln(E_i)-ln\,\mu_E\right)^2\right]}{-\left(N\mu_E\sqrt{2\pi}\right)(\sigma_E)^N\,\Gamma(-N+2)}\right] \tag{3.5.1.30}$$

Maximizing the log-posterior σ_E

$$\frac{\partial}{\partial\sigma_E}\left[\begin{array}{l}log(c)+log[2]-\sigma_E+log\left[(\sigma_E)^{-N}\right]+\\ \qquad\qquad\left[-\frac{1}{2(\sigma_E)^2}\sum_{i=1}^{N}\left(ln(E_i)-ln\,\mu_E\right)^2\right]\end{array}\right]=0 \tag{3.5.1.31}$$

where: $c=\left[-\left(N\mu_E\sqrt{2\pi}\right)(\sigma_E)^N\,\Gamma(-N+2)\right]^{-1}$

The most likely value for σ_E

$$\hat{\sigma}_E=\sqrt{\frac{3}{N}\sum_{i=1}^{N}\left(ln(E_i)-ln\,\mu_E\right)^2} \tag{3.5.1.32}$$

Maximizing the log-posterior for μ_E

$$\frac{\partial}{\partial\mu_E}\left[\begin{array}{l}log(c)+log[2]-\sigma_E+log\left[(\sigma_E)^{-N}\right]+\\ \qquad\qquad\left[-\frac{1}{2(\sigma_E)^2}\sum_{i=1}^{N}\left(ln(E_i)-ln\,\mu_E\right)^2\right]\end{array}\right]=0 \tag{3.5.1.33}$$

The most likely value for μ_E

$$\frac{1}{\mu_E(\sigma_E)^2}\sum_{i=1}^{N}\left(ln(E_i)-ln\,\mu_E\right)=0 \tag{3.5.1.33}$$

$$ln(\mu_E)=\frac{\sum_{i=1}^{N}ln(E_i)}{N} \tag{3.5.1.34}$$

$$\hat{\mu}_E=\sqrt[N]{\sum_{i=1}^{N}(E_i)} \tag{3.5.1.35}$$

3.5.2 Percent Accuracy of Experts

The second metric of expertise is the percent accuracy of independent experts employed in the meta-analysis. It is defined as overall accuracy or the probability of expert(s) correctly predicting all states (i.e. positive and negative, present and absent, yes and no, etc.) of nature. Percent accuracy as used in this research is illustrated in the Table 2. Of the 80 data points collected (Table 3), eleven were given in the form of sensitivity and specificity only. In these cases, the mean of the sensitivity and the specificity values were used as estimates of the overall accuracy. Sensitivity, as demonstrated in our example, is the ratio of the true-positive judgments and all positive states, or probability that the expert's judgment is positive when the true state of nature is positive. Specificity, on the other hand, is ratio of true-negative judgments and all negative states, or the probability that the expert's judgment is negative when the true state of nature is negative.

Table 2. Illustration of Measure of Expertise

		True State of Nature/ Event	
		True State is Positive (+)	**True State is Negative (-)**
Expert Judgment about the True State of Nature	**Expert Predicts Positive (+)**	**<u>A</u>** (True Positives)	**<u>B</u>** (False Positives)
	Expert Predicts Negative (-)	**<u>C</u>** (False Negatives)	**<u>D</u>** (True Negatives)

Figures 3 and 4 illustrate the 80 data points expressed as relative frequencies of percent accuracies in a histogram. Furthermore, the histogram was subsequently evaluated for characteristics of various probability density functions; the beta and truncated normal distributions were the best fits. The resulting beta fit (Figure 3) implies that the probability of expert percent accuracy being at and near 100 sharply decreases towards zero. In contrast, percent accuracy is expressed in Figure 4 by a normal distribution, truncated on the right at 100. The truncated normal distribution shows that experts are reasonably likely to obtain 100% performance accuracy. The mean of this distribution is located at 74.7%. The alpha 1 and alpha 2 parameters of the beta distribution in Figure 3, and the mean and standard deviation parameters of the truncated normal distribution in Figure 4 are used to develop likelihood functions in Chapters 4 and 5.

Table 3. Meta-Analysis Results of Interdisciplinary Percent Accuracies

Percent accuracies							
1	23	**21**	74	**41**	83.7	**61**	99
2	40	**22**	77.4	**42**	89	**62**	98
3	48	**23**	75	**43**	80.6	**63**	94
4	49.5	**24**	79.5	**44**	89.8	**64**	90.5
5	50	**25**	72.8	**45**	87	**65**	96
6	41	**26**	80	**46**	88.7	**66**	76.2
7	50	**27**	71	**47**	81.5	**67**	73.5
8	42	**28**	76	**48**	80.5	**68**	73
9	52.9	**29**	73	**49**	89	**69**	85.8
10	52.6	**30**	76	**50**	90	**70**	93
11	52.3	**31**	71	**51**	84.6	**71**	98.8
12	57	**22**	80	**52**	82	**72**	70.2
13	56.9	**33**	72.1	**53**	85	**73**	71
14	56	**34**	77	**54**	86	**74**	97
15	55.7	**35**	73	**55**	89.5	**75**	68.4
16	51	**36**	78	**56**	83	**76**	74.5
17	57	**37**	80	**57**	87	**77**	91.9
18	69.5	**38**	72	**58**	84.3	**78**	68
19	70.3	**39**	80	**59**	96.8	**79**	74
20	67	**40**	76	**60**	96	**80**	96

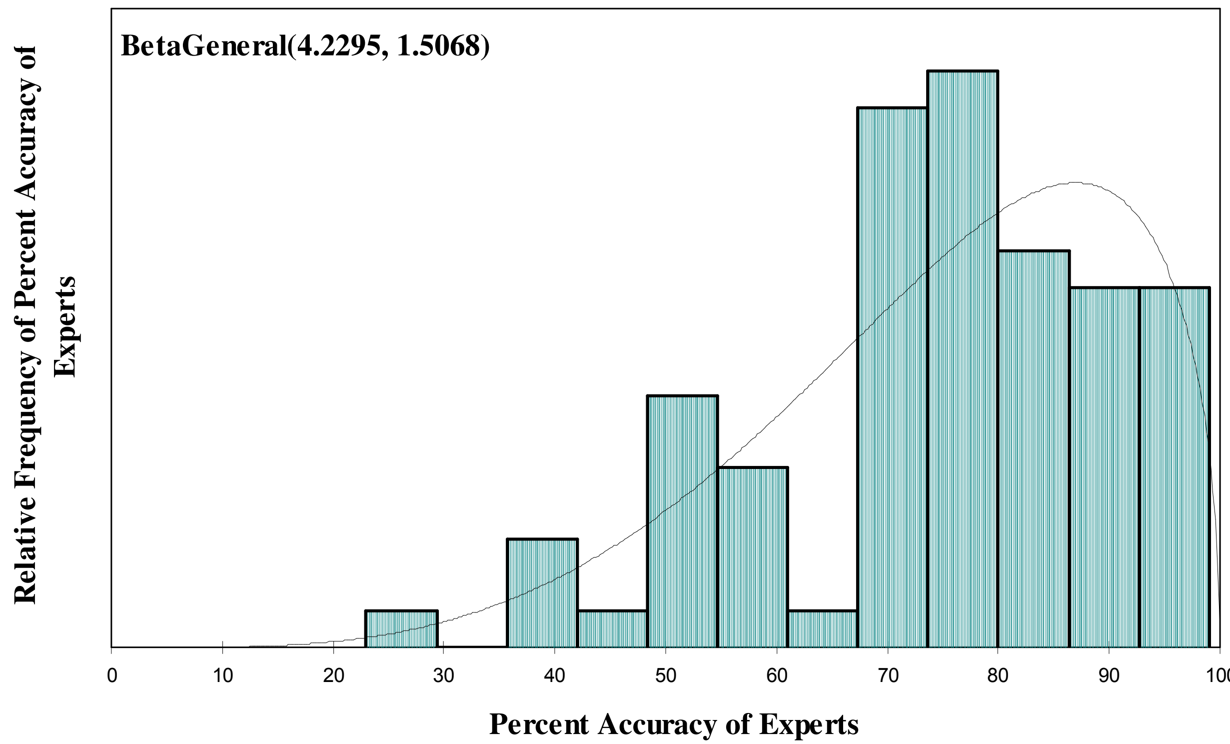

Figure 3. Beta Probability Density Function Fitted to Relative Frequency of the Percent Accuracy Values for Experts Obtained From Meta-Analysis

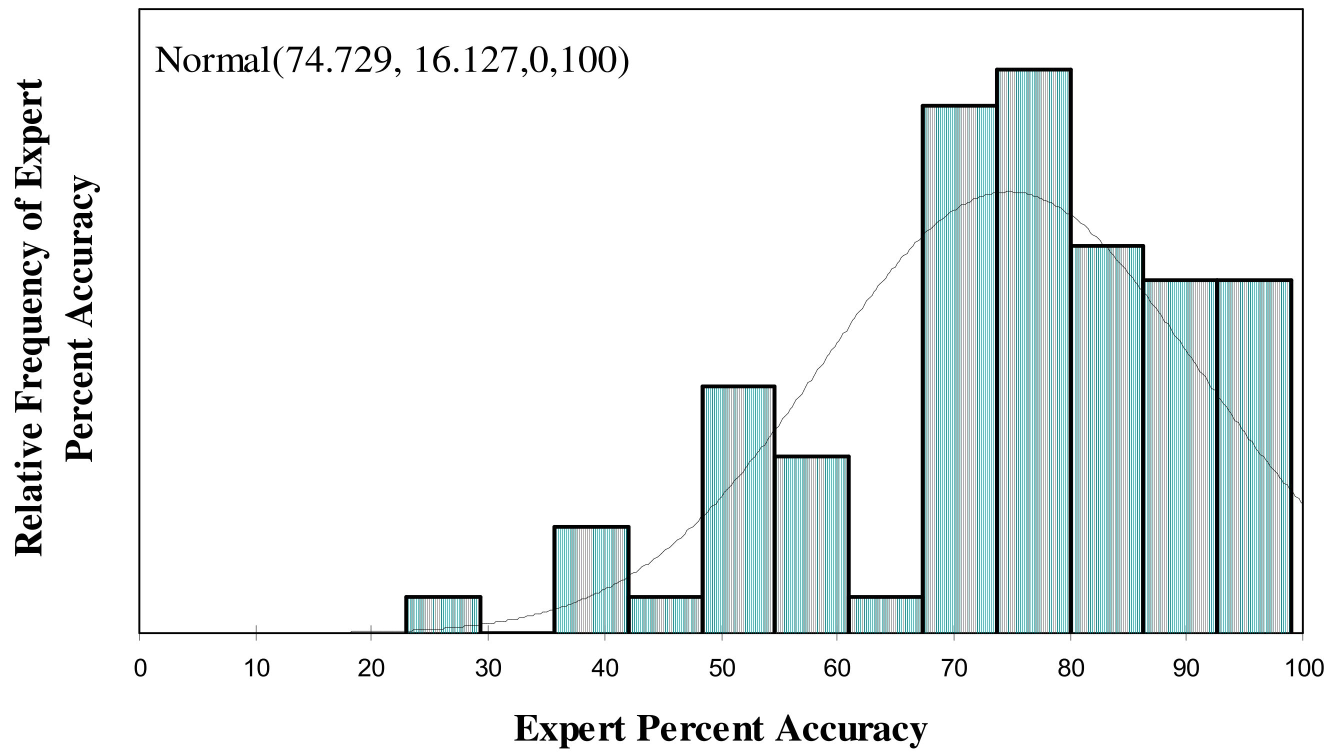

Figure 4. Truncated Normal Probability Density Function Fitted to Relative Frequency of the Percent Accuracy Values for Experts Obtained From Meta-Analysis.

3.5.3 Descriptive Characteristics of Attributes

Table 4 contains eight of the nine categorical attributes along with their respective mean expert accuracies. The table reports that expert who are certified in expertise, from companies who specialize in the elicited issue or who have attained specialized expertise beyond their terminal degrees are more likely to be accurate than those who do not. In addition, results in Table 4 also suggest an ordinal relationship between the average level of formal education to expert performance.

Analysis of the data showed a natural grouping of accuracy below and above 70%. As a result, each attribute was evaluated for correlation with accuracy, dichotomously organized into subcategories of 0-69.9% and 70-100%. The Eta correlation coefficient was used to represent the strength of the relationship between expert accuracy (dependent variable) and each attribute (independent variable). Eta correlation, also known as the coefficient of non-linear correlation, is a ratio of the partial sum of squares of each independent variable and the total sum of squares. Furthermore, it describes a curvilinear relationship between dependent and independent variables, and is comparable to R-squared.

Tables 7 and 8 shows the resulting correlation strengths of each attribute to accuracy. Among the attributes investigated, the following were significantly correlated with accuracy: certified or specialized training in expertise; level of educational achievement; publication(s) in expertise or general field; and the type of institution. The finding of significant correlation between experience and performance has been documented by others. Vegelin et al (2003) concluded that experience significantly influences accuracy. Sorrento and Pichichero (2001) also surmised that “experience is an excellent teacher and this may have contributed to the greater accuracy”.

Table 4 also shows that experts who are certified or recipients of specialized training in elicited area are 17% more accurate than those who were not. Similarly, the accuracy of experts with post-doctoral training or medical doctors with specialized training was higher than undergraduate, graduates, and PhDs and MD without specialized training. Experts who are published on the elicited topic or in the field performed significantly better (~10%) than others. In addition, the data showed experts from government and private organizations performing better than those from academia. A possible explanation could be the access to resources. Other attributes listed in Table 4, such as nomination by peers; membership in professional organizations; company/organization specialization in specific/similar topic; and frequency of event/disease, revealed weak correlations with accuracy. In fact, membership in professional organizations and peer nominations, appeared to be almost uncorrelated (Eta=0.031) and (Eta=0.089), respectively.

Table 4. Mean Accuracy of Experts with Subsequent Attributes

ATTRIBUTES	ACCURACY	
	Mean	**Std. Dev.**
Nominated by peers as expert in field(s) or expertise	-----	-----
• Yes	76.8	15.9
• No	73.1	16.3
Certified or received specialized training in expertise	-----	-----
• Yes	80.0	13.3
• No	63.1	15.9
Publication(s) in expertise or general field	-----	-----
• Yes	77.8	14.8
• No	67.4	17.1
Member of professional organization in expertise/field	-----	-----
• Yes	75.2	16.5
• No	70.6	12.6
Company/organization specializes in specific/similar topic	-----	-----
• Yes	80.8	13.7
• No	72.8	16.5
Type of Institution	-----	-----
• Private	78.7	10.0
• Academic	71.8	16.4
• Private & Public	74.4	15.2
• Government	79.7	13.3
Average level of formal education	-----	-----
• Undergraduate	68.9	21.6
• Graduate Student/ MD-Intern	65.6	16.7
• PhD/ MD	65.0	11.9
• Post-Doc/ MD-Specialist	81.0	11.6
Frequency of event/disease	-----	-----
• Rare	72.8	16.9
• Moderate	74.7	17.2
• Frequent/ Common	75.4	15.7

Tables 5 and 6 are extensions of Table 4. Both tables address descriptive characteristics of the "average years of practical experience in expertise" attribute. This attribute is divided into five subcategories. The subcategory of "None" typically represents experts who are elicited for issues not explicitly related to their expertise. For example, experts are sometimes asked to give their judgments about the future of newly developed technologies or fields. This lack of experience concerning the information being elicited is represented by "None" in the data set.

Table 5 demonstrates that experts with five to less than ten years, or with greater than ten years experience had similar accuracy scores, and their accuracy was significantly higher than experts who had less than five years experience. In addition, experts with ten or more years of experience were slightly less accurate than those in the previous subcategory. This implies that experts with the most years of experience are not necessarily the best suited experts. However, this conclusion cannot be definitively stated because of the limited data availability for this attribute. The authors of most expert judgment case studies (67.5%) did not state the experts' years of experience. Many qualitatively expressed experience of experts in the following forms: "specialist … with special expertise in" (Afset et al. 1996); "considerable experience" (Jankovic 2000); "well-trained" (Sboner 2004); and "each was experienced in" (Bruynesteyn et al. 2002). These expressions are very subjective and were codified into the "Not Stated" category.

Table 5. Average years of practical experience in expertise

	ACCURACY		DISTRIBUTION OF CASE STUDIES
	Mean	Std. Dev.	
None	39.350	23.1224	2.5%
> 0 to < 5 years	60.791	13.8462	13.75%
5 to < 10 years	68.271	16.5186	8.75%
≥ 10 years	65.271	18.4848	7.5%
Not Stated	80.760	11.7453	67.5%

Table 6. Correlation of Average years of experience and accuracy under three options

Options	Eta Correlation*
• Not Stated case studies separate as in Table 2	0.293
• Not Stated case studies are included in the 5 to <10 years category	0.484
• Not Stated case studies are included in the ≥10 years category	0.506

*Average years of experience is evaluated for correlation with accuracy from 0-69.9% and 70-100%

With the "Not Stated" case studies excluded from the analysis, the correlation of the "average years of practical experience in expertise" attribute to accuracy was 0.293 (Table 6). Table 6 listed two other correlation relationships, based on possible quantitative implications of these qualitative expressions. In one option, the "Not Stated" case studies were placed into the "5 to <10 years" category. This option yielded a stronger correlation value of 0.484 and a still stronger correlation value of 0.506 after the "Not Stated" case studies were added to the "≥10 years" category. The assumptions made are based in part on studies with both the subjective and quantitative expressions on experience. Brown et al. (2004) for example, expressed "very experienced" as "an average of 17.5+/-11.5 years of total … experience".

3.5.3 Logarithmic Relationship Curves

All performance values or data points used to formulate the distributions in Figures 3 and 4 are innately associated with an array of attributes. The results of this meta-analysis identified logarithmic relationships of attributes to performance or accuracy (Figures 5, 6 and 7). Figures 5 and 6 graphically illustrate the logarithmic relationship of cumulative attributes to performance. This implies performance *"p"* is equivalent to sum of a task-related constant *"T"*, and the product of the constant β and the summation of the attributes $\sum_{j=1}^{N} x_j$:

$$p = \beta \, ln\left(\sum_{j=1}^{N} x_j\right) + T = \beta \, ln(X_i) + T$$

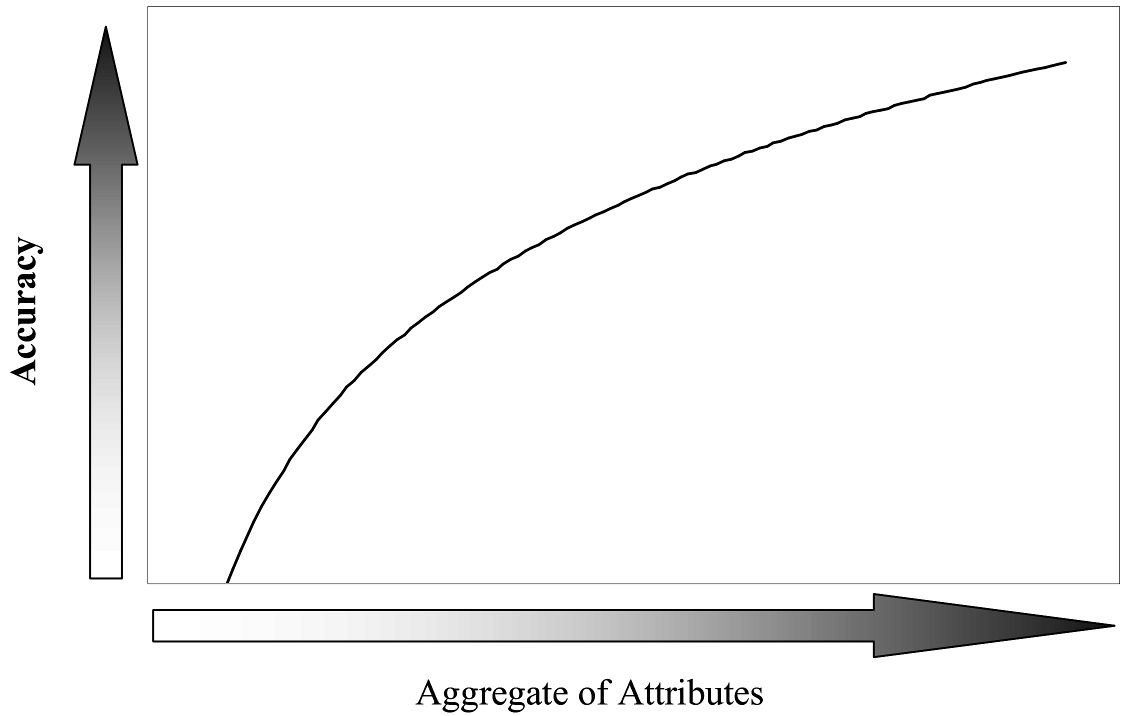

Figure 5. Accuracy vs. Logarithmic Aggregated Expert Attributes

$$p = \beta \, ln\left(\sum_{j=1}^{N} x_j\right) + T$$

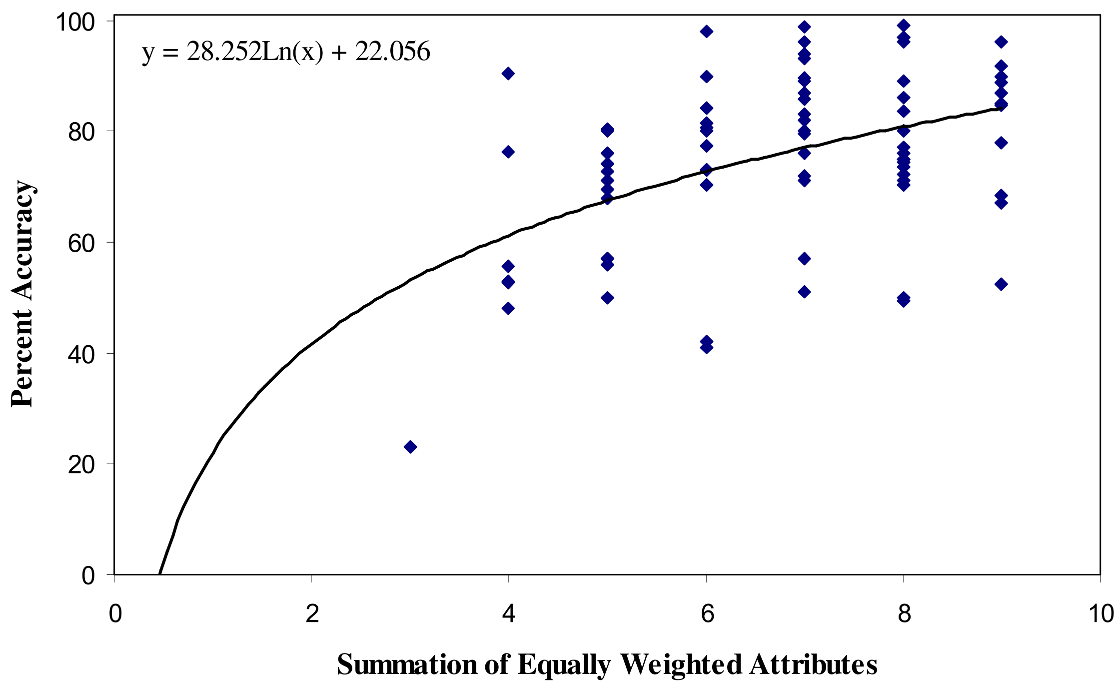

Figure 6. Accuracy vs. Non-Weighted Aggregated Expert Attributes

Each performance score (percent accuracy) is associated with an array of attributes $x_1,...x_N$. The attributes are associated with two or more nominal or ordinal states. For example, the attribute "Certified or received specialized training in expertise" has two nominal states. If the expert has this attribute the state is "YES", and if the expert does not, the state is "NO". Another example is the attribute, "Average level of formal education". This attribute has the following multiple ordinal states: Undergraduate, Graduate Student/ MD-Intern, PhD/ MD, and Post-Doc/ MD-Specialist. Ordinal states differ from nominal by the intrinsic hierarchy.

Tables 7 and 8 present a taxonomy for the states of common attributes and their subsequent correlation to performance (accuracy). The nominal and ordinal states of each attributes are assigned empirical values. The taxonomy also includes coefficient values that characterize the strengths of correlation between each attribute and accuracy.

Using the values within tables 7 and 8, the mathematical and graphical relationship between attributes and accuracy was refined. Figures 7 and 8 demonstrate the effect of the correlation coefficients c_{ij}. Figure 7 displays the fit curve the following logarithmic curve, to the data:

$$p_i = \beta \, ln\left(\sum_{j=1}^{N} x_{ij} c_{ij}\right) + T_i$$

A comparison of the fits in both Figures 6 and 8 shows, the data is better fitted to the curve with the inclusion of the correlation coefficients. This means expert performance or accuracy is better predicted using the mathematical relationship in Figure 8. Validation of these fits occurred through two case studies, the results of which are presented in Chapter 6.

Table 7. Taxonomy of the States and the Correlation Coefficients of Attributes (contd.)

ATTRIBUTES x_{ij}	State of Attribute $x_{ij} = 0,1$	Correlation Coefficient c_{ij}
Nominated by peers as expert in field(s) or expertise	-----	0.089
• No	0	
• Yes	1	
Certified or received specialized training in expertise	-----	0.441
• No	0	
• Yes	1	
Publication(s) in expertise or general field	-----	0.260
• No	0	
• Yes	1	
Member of professional organization in expertise/field	-----	0.031
• No	0	
• Yes	1	
Company/organization specializes in specific/similar topic	-----	0.146
• No	0	
• Yes	1	

Table 8. Taxonomy of the States and the Correlation Coefficients of Attributes

ATTRIBUTES x_{ij}	State of Attribute $x_{ij} = 0,1$	Correlation Coefficient c_{ij}
Academic Institution	1	0.263
Government/Private Institution	1	0.263
Average level of formal education	-----	-----
• Undergraduate	1	0.059
• Graduate Student/ MD-Intern	1	0.324
• PhD/ MD	1	0.234
• Post-Doc/ MD-Specialist	1	0.410
Average years of practical experience in expertise	-----	------
• > 0 to < 5 years	1	0.325
• 5 to < 10 years	1	0.164
• ≥ 10 years	1	0.084
Frequency of event (i.e. disease)	-----	------
• Rare	1	0.064
• Moderate	1	0.069
• Frequent	1	0.110

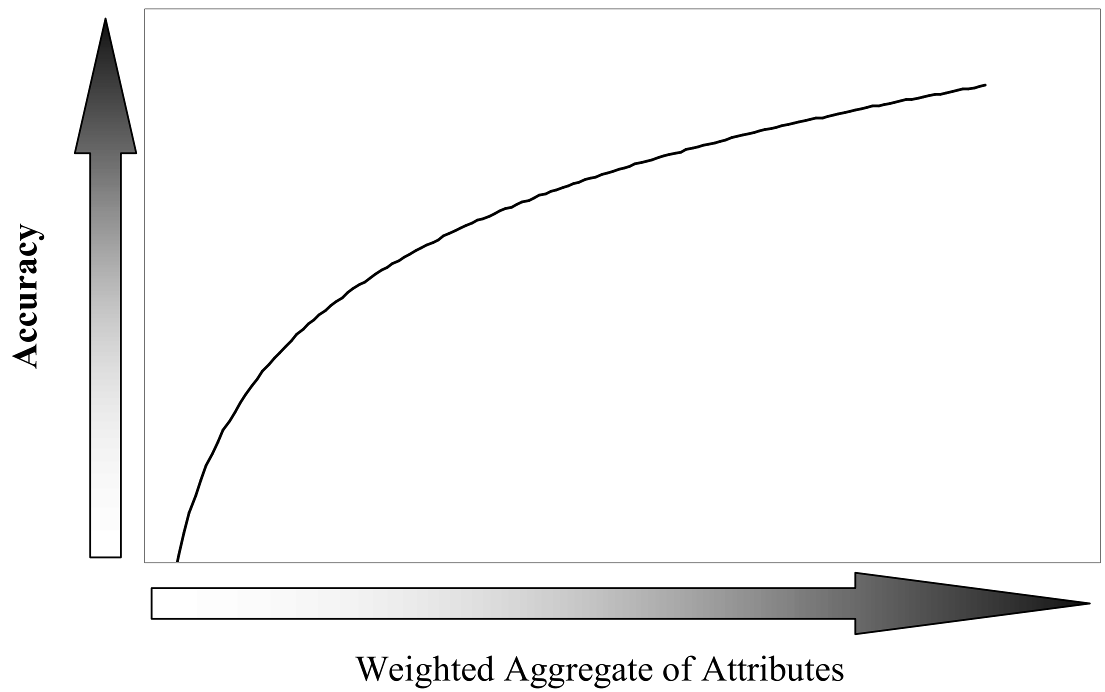

Figure 7. Accuracy vs. Logarithmic Aggregated Expert Attributes

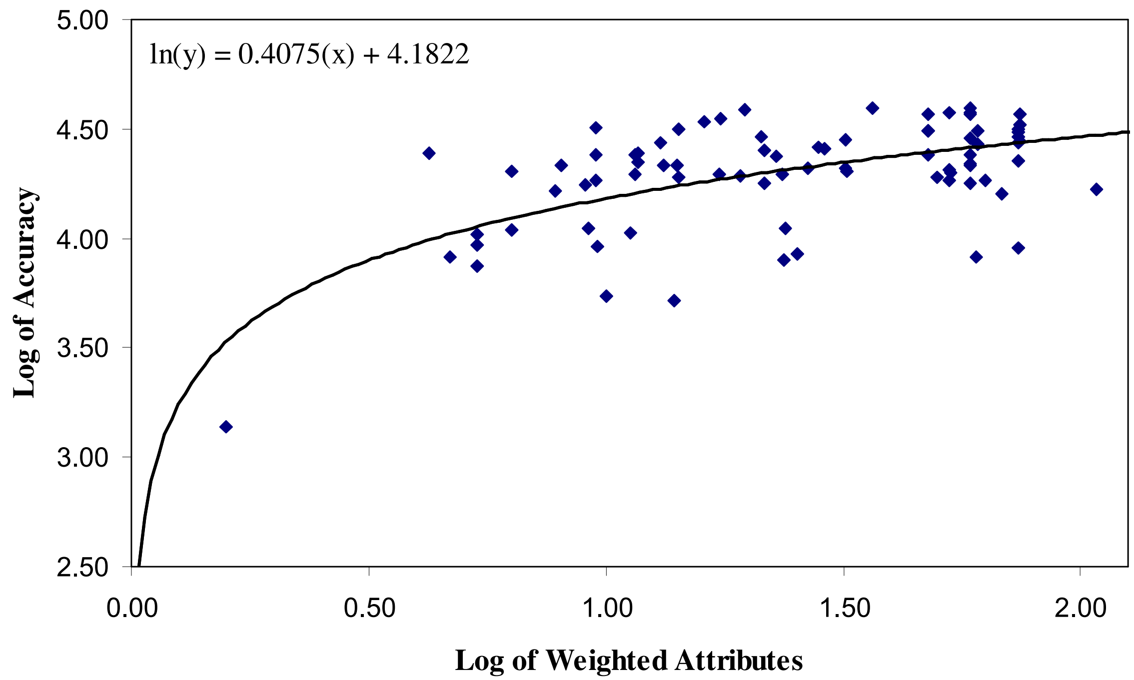

Figure 8. Accuracy vs. Logarithmic Aggregated Expert Attributes

3.6 SUMMARY

Validated expert judgment case studies were evaluated for two common measures of expertise, mean absolute percentage error and percent accuracy of experts. The meta-analysis of case studies yielded 58 data points for mean absolute percentage error, and 80 data points for percentage accuracy. Both data sets represented in the histograms of Figures 1 through 4 depicted the relative frequency of individual data in each set. The resulting histograms were then fitted to various probability density functions. The exponential and lognormal functions were consistent with the histogram denoting mean distribution errors of 12.1% and 14.9%, respectively. Similarly, the beta and a normal distribution truncated on the right were consistent with the histogram illustrating percent accuracy of experts.

Expert performance or percent accuracy was evaluated for correlation with the attributes of experts. Each attribute was evaluated for correlation with the accuracy subcategories of 0-69.9% and 70-100%. The data in this study suggest that certified or specialized training in expertise; level of educational achievement; publication(s) in expertise or general field; the type of institution, and average years of experience are reasonable predictors of accuracy. In addition, membership in professional organizations and peer nominations, were the least correlated, with correlation coefficients 0.031 and 0.089, respectively.

The performance of experts is associated with an array of attributes. The results of this meta-analysis identified logarithmic relationships of attributes arrays to performance or accuracy (Figures 5 through 8). Figures 5 and 6 graphically illustrate the logarithmic relationship of cumulative attributes to performance. While Figures 7 and 8 demonstrate a refined logarithmic relationship by incorporating the correlation coefficients c_{ij} from the taxonomy in Tables 7 and 8. The comparison of fits in both reveal the data is better fitted to the curve in Figures 7 and 8.

The author believes the published studies of validated expert judgment used in this research are an appropriate representation of the available studies. However, the author also recognizes the deficiencies in the limited descriptions or inclusion of expert background or attributes in published case studies. These deficiencies may have reduced the strength of the correlation between attributes and accuracy. As a result, distribution of correlation coefficients that represent associated uncertainties is recommended.

CHAPTER 4

BAYESIAN ASSESSMENT OF TRUE VALUES FOR CONTINUOUS QUANTITIES

4.1 ABSTRACT

Experts are often elicited to make quantitative judgments about continuous quantities. Decision makers and other stakeholders are interested in the deviation of expert estimates from the true values of known and unknown quantities. There exist several studies with methodologies to calibrate experts or predict expert performance. However, these methodologies require historical knowledge of an expert's performance. Other studies present models whose applications are limited to specific case studies, and narrow scopes within a discipline or field. This chapter presents models within the Bayesian framework to predict or forecast the true values of unknown quantities based on the input of expert estimates.

4.2 INTRODUCTION

In the absence of sufficient data, the decision maker lacks perfect knowledge about quantities of interest, whose true values are unknown. As a result, the elicitation of expert opinion or judgment regarding these unknown quantities of interest is employed. Experts possess unique or specialized knowledge or skill(s) about the quantities of interest to the decision maker. Experts are elicited for both quantitative and qualitative judgments regarding select quantities. Quantitative judgments are given in the form of point estimates, quartiles, and distributions. However, this dissertation focuses on the point estimates of experts. The quantities of interest to the decision makers may be categorized as continuous or categorical. Within a range, continuous quantities have infinite possible values. Examples of continuous quantities are length, time, weight, stress, strain, as well as volume, temperature, humidity, failure rates, and dosage. Categorical quantities, on the other hand, have a finite number of possible values within a range. Examples of categorical quantities include, condition present or absent, correct or incorrect diagnosis, and selection of specific condition. Computations in this chapter are based on data of continuous quantities; categorical quantities are covered in the next chapter.

A unique property of continuous quantities or variables is enabling the determination of the deviation of expert estimates from the true value. Experts frequently provide estimates that are some measure from the true value of the quantity. This is exemplified in expert estimates of temperature, in financial forecasts, of appropriate dosages, and of time to failure. This chapter is devoted to the development of the likelihood functions and posterior distributions for the true value of an unknown quantity given the estimate of an expert or multiple experts. In addition, the calibration equations for determining the unknown quantities are also presented. The resulting equations are significant to decision makers and stakeholders, who desire to know the approximate value of an unknown quantity or the error of an expert's judgment.

4.3 METHODOLOGIES FOR CALIBRATING EXPERT JUDGMENT

Literature contains an abundance of resources to estimate the performance of experts and their estimates. Some authors recommend various models to represent expert judgment and others suggest appropriate methodologies for aggregation of judgments. However, scientists are more reserved in advocating for specific models, with empirically defined parameters to represent expert performance and estimates. The reluctance stems primarily from the view that a universal equation or sets of equations cannot reasonably predict the true values of all quantities. As a result, scientists have traditionally focused their attentions on the calibration of expert judgment.

4.3.1 Cochran, Weiss, and Shanteau (CWS) Index of Expertise

"When it is clear that an outcome measure captures expertise, it is appropriate to use it as a means to identify the expert" (Weiss and Shanteau 2003). Cochran, Weiss and Shanteau (2003) argued that evaluative skill is the basic cognitive ability that characterizes all these areas of expertise. As a result, they created the CWS index of expertise, ($CWS\ index = \dfrac{discrimination}{inconsistency}$), a ratio of discrimination over inconsistency. "Discrimination refers to the judge's differential evaluation of the various stimuli similarly over time. Consistency refers to the expert evaluation of the same stimuli similarly over time; inconsistency is its complement" (Weiss and Shanteau 2003).

4.3.2 Bayesian Calibration Model: Debiasing Expert Overconfidence

Clemen and Lichtendahl (2002) proposed approaches for modeling and debiasing expert confidence. The model demonstrated the ability to debias expert probabilities, based on past performance data (previous assessments and realizations for a number of uncertain variables), and the use of Bayesian methods to update model parameters in the prior distributions. Both sing-expert model and multiple-expert hierarchical model were developed (Clemen and Lichtendahl 2002).

4.3.3 Calibration of Expert Judgment per Fractile

Calibration can be measured empirically in experiments that involve many assessments of quantities about which assessors have some relevant but imperfect knowledge, and whose true value can be found by the experimenter. Typically in these experiments subjects are asked about their degree of uncertainty about such things as the populations of countries, dates of historical events, or meanings of words, which can be easily verified.

To measure calibration for a set of assessments of discrete probabilities, they are partitioned into subsets with the same or similar assessed probabilities. These probabilities are then plotted against the actual faction of each subset that judge the curve and should be near the diagonal. For an under-confident judge, the assessed probabilities are nearer 0.5 than they should be; more typically, judges are overconfident and the probabilities are assessed to near certainty (0 or 1). An analogous calibration curve can be compiled for assessments of continuous distributions for unknown quantities. For each item the fractile of the assessed distribution at which the true value occurs is recorded. These fractiles form a distribution of values between 0 and 1, and the cumulative of this distribution is also a calibration curve. Again, the curve of a well-calibrated assessor would be the diagonal. Two commonly used measures of calibration on unknown quantities compare the predicted probability of falling within a particular interval with the actual number of values inside it (Henrion and Morgan, 1990, Pg110-112)

All calibration methodologies previously discussed require individual historical expert performance data, which are often not available. The methodology developed in this chapter is independent of the previous performance of a particular expert. It is conditional upon the estimates of the expert, a condition which imitates reality.

4.4 METHODOLOGY

An extensive meta-analysis of expert judgment validation literature was conducted to provide values for essential model parameters within the Bayesian framework. Meta-analysis is the process of performing analysis on amalgamated results of various published studies into a common metric. The search for relevant publications entailed a general survey of past and current literature, internet publications, refereed and non-referred sources of articles on *expert opinion* or *expert judgment* accuracy.

The wide literature search included in part the following databases: WorldCat, Ingenta, INSPEC, Applied Science and Technology, LexisNexis Academic, SourceOECD and Pubmed, covering fields such as: financial forecasting, weather forecasting, medicine, human nutrition, and several engineering disciplines. The search generated over 850 abstracts, from which approximately 112 articles were initially selected, and 58 found suitable for inclusion in the data analysis.

Percent error or data facilitating the calculation of absolute percent error (APE) were used as the measure of expertise by which to evaluate and select each article. Suitable articles were selected based on their inclusion of this data. The data were fitted to various probability distribution functions, and the most appropriate fits were chosen, for use in the Bayesian framework.

4.5 EXPERT PERFORMANCE CALIBRATION IN A BAYESIAN FRAMEWORK

The Bayesian framework is routinely used to aggregate and represent the uncertainty in expert judgment. The Bayesian paradigm (Bayes' Theorem) combines prior information with new information of an unknown quantity to represent the current state of the quantity. As new information becomes available, the current state of knowledge regarding the unknown quantity is changed. Bayes' Theorem was named after, the 18th century mathematician and cleric, Reverend Thomas Bayes, who derived a special case of this theorem. Initial formal applications of Bayesian theory to subjective probability and utility was developed by deFinetti in 1930 and Ramsey in 1931 (Broemeling, pg 41, 1985).

Bayesian methodologies differ from and have several advantages over classical inferential or statistical models. Perhaps the most striking characteristics are the ability of Bayesian models to formally incorporate prior knowledge and subjective data into the computations. Prior knowledge could take the form of common continuous and discrete probability distributions. These distributions include the non-informative uniform, the exponential, the binomial, the normal and lognormal, as well as the weibull, the beta, and the pert. In addition to objective data, Bayesian models also allow for the incorporation of expert judgment or other subjective information, as will be demonstrated in this chapter. Furthermore, with limited or scare data, Bayesian models make similar inferences to those of classical inferential or statistical models, and can accommodate data in any order.

Conversely, Bayesian methodologies are associated with some disadvantages. Appropriate selection of prior and likelihood distribution pairs can become problematic. The normalization factor for non-conjugate prior and likelihood pair are often complex and difficult to solve. Furthermore, prior distributions are prone to misuse, by their prejudicial selection to manipulate posterior distributions or results.

4.5.1 Bayes Theorem Overview

Bayes's Theorem is a mathematical methodology used for the calculation of conditional probabilities. Conditional probabilities denote the degree of belief in a proposition based on assumption(s) that another argument is true. The following depicts the general equation of Bayes' Theorem for evaluating an unknown quantity *"x"*:

$$\pi(x|x') = \frac{L(x'|x)\pi_o(x)}{\int L(x'|x)\pi_o(x)dx} = k^{-1}L(x'|x)\pi_o(x) \qquad (4.5.1.1)$$

where:

$\pi(x|x')$ is posterior distribution representing the decision maker's posterior state of knowledge about the unknown quantity, x, given that he has received the set of the experts' opinions x',

$L(x'|x)$ is the likelihood of the evidence x' given that the true value of the unknown quantity is x.

$\pi_o(x)$ is the decision maker's prior or initial state of knowledge about the unknown quantity x (prior to receiving the opinion of the experts)

k is a normalization factor that makes $\pi(x|x')$ a probability distribution:

$$k = \int L(x'|x)\pi_o(x)dx$$

A critical component of this theorem is the likelihood function, $L(x'|x)$, which reflects the probability that a particular population(s) would produce a particular value(s). The likelihood function is a conditional joint probability function. In the case of multiple data points or multiple experts (in this chapter), the likelihood function is created by product of the probability distribution function $L(x'|x)$, of each of the data point. The general mathematical representation of the likelihood function for multiple data points is below:

$$L(x_1,\ldots x_N|x) = L(x_1|x)\ldots L(x_N|x) = \prod_{i=1}^{N} L(x_i|x) \qquad (4.5.1.2)$$

Generally, the individual likelihood functions of expert assessments of various quantities differ. However, the likelihood functions developed in this chapter considers experts of all discipline as one unit, and "should be closely related to the calibration measures for that expert" (Mosleh 1981, pg22).

4.5.2 Likelihood Function for Expert Evidence

There exist numerous suppositions of appropriate functions to depict the likelihood of expert evidence, given that the true value of an unknown quantity is *"x"*. These functions include in part, common probability distributions such as exponential, normal, lognormal, as well as weibull, beta, and gamma. Selection of appropriate distributions and evaluation of model parameters representing expert evidence is generally subjectively decided. Scientists select models and evaluate parameters based on limited or no empirical evidence, or theoretical presumptions. Few studies have been performed to determine the true empirical value(s) of parameter(s) in models that evaluate expert judgment on a specific issue, within a discipline, or across disciplines.

This section is devoted to the development and validation of the likelihood functions for the true value of an unknown quantity given the estimate of an expert or multiple experts. Likelihood functions and model parameters of continuous quantities are constructed and derived from empirical results of the meta-analysis of literature. The meta-analysis across multiple disciplines reveals that the exponential and lognormal distributions appropriately depict the likelihood of expert evidence, given a fixed value for the unknown quantity. To our knowledge, this work contains the most comprehensive set of validated expert judgment case studies, used in the Bayesian framework.

All validated expert judgment case studies employed in the meta-analysis, were evaluated for a common metric of expertise. The common metric, absolute percentage error, APE or E, was either implicitly stated or extracted from data in each case study. Absolute percentage error is the ratio of the absolute difference of estimated and actual value over actual value.

$$APE = E = \left|\frac{u' - u}{u}\right| \cdot 100 \qquad (4.5.2.1)$$

From the relationship of u, u' and E in Equation 4.5.2.1, the equation for the true value of the unknown quantity can be stated as follows:

$$u = \frac{u'}{1 + 0.01|E|} \qquad (4.5.2.2a)$$

or

$$u = \frac{u'}{1 \pm 0.01 \cdot E} \tag{4.5.2.2b}$$

The variable E denotes a probability density function of absolute percentage errors $(E_1 ... E_N)$, and is derived from the estimates of experts about elicited quantities along with the true values the elicited quantities. Therefore, the resulting variable u is a distributed quantity, such that given an expert's single estimate u' about an elicited quantity, there exist a distribution of possible true values u.

<u>Likelihood Function for a Single Expert: Exponential Distribution Model</u>

The function $L(u'_i\, |u)$ denotes the likelihood of the expert *i's* estimate, u'_i, given that the true value of the unknown quantity of interest to the decision maker is, *"u"*. This likelihood function in combination with a prior distribution is used, to determine the posterior state or Bayesian posterior distribution of *"u"*:

$$\pi(u|u'_i) = \frac{L(u'_i\, |u)\pi_o(u)}{\int L(u'_i\, |u)\pi_o(u)du} = k^{-1} L(u'_i\, |u)\pi_o(u) \tag{4.5.2.3}$$

where:

$\pi(u|u'_i)$ represents the posterior distribution of the decision maker's state of knowledge about the unknown quantity *u*, given that he or she has received evidence or estimate u'_i from expert *i*

$\pi_o(u)$ represents the decision maker's prior state of knowledge about the true value of the unknown quantity *u*

In constructing the likelihood function, the function of absolute percentage error, $f(E)$, must be developed. The function $f(E)$ is derived below from the relationship of u and all u'_i s. The following is a general relationship of the true value of an unknown quantity to the expert estimate and error.

From Equation 4.5.2.2:

$$d(u') = d(u(1 + 0.01 \cdot E)) \tag{4.5.2.4}$$

and

$$du' = u \cdot dE \qquad \Rightarrow \qquad \frac{dE}{du'} = \frac{100}{u} \tag{4.5.2.5}$$

and

$$f(u')du = f(E)dE \tag{4.5.2.6}$$

and

$$f(u')du' = f(E)\frac{dE}{\mathrm{du'}} \qquad \Rightarrow \qquad f(u') = \frac{100}{u} f(E) \tag{4.5.2.7}$$

Equations 4.5.2.2 and 4.5.2.4 thru 4.5.2.7 indicate that the random variable u' is a function of E. As a result, the likelihood function $L(u'_i|u)$ can be restated as $L(E_i|u)$ and the general Bayesian posterior distribution in Equation 4.5.2.3 as:

$$\pi(u|E_i) = \frac{L(E_i|u)\pi_o(u)}{\int L(E_i|u)\pi_o(u)du} = k^{-1}L(E_i|u)\pi_o(u) \tag{4.5.2.8}$$

The exponential error distribution $f(E)$ was experimentally attained in Chapter 3, and contains the parameters α and β, wherein:

$$f(E_i) = \alpha e^{-\frac{1}{\beta}|E_i|} \tag{4.5.2.9}$$

Equation 4.5.2.9 is a two-parameter exponential probability density function which contains a random sample of independent expert absolute percentage errors $(E_1 \ldots E_N)$.

Assuming a normalized distribution wherein $\int f(E_i)dE_i = 1$, then $\alpha = \frac{1}{\beta}$, and Equation 4.5.2.9 reduces to:

$$f(E_i) = \frac{1}{\beta} e^{-\frac{1}{\beta}|E_i|} \tag{4.5.2.10}$$

The likelihood of the absolute error E_i of expert "i" given that the true value of the elicited quantity is "u" is:

$$L(E_i|u) = \frac{100}{u} \cdot \frac{1}{\beta} e^{-\frac{1}{\beta}|E_i|} \tag{4.5.2.11}$$

or

$$L(u'_i|u) = \frac{100}{u} \cdot \frac{1}{\beta} e^{-\frac{1}{\beta}\left|\frac{u'_i - u}{u}\right|} \tag{4.5.2.12}$$

Likelihood Function for Multiple Experts: Exponential Distribution Model

Similarly, the likelihood function for multiple experts ($1 \ldots N$) is the product of the likelihood function for each expert, such that:

$$L(u'_1 \ldots u'_N|u) = \prod_{i=1}^{N} \left(\frac{100}{u\beta} e^{-\frac{100}{\beta} \cdot \left|\frac{u'_i - u}{u}\right|} \right) \tag{4.5.2.13}$$

or

$$L(u'_1 \ldots u'_N|u) = \left(\frac{100}{u\beta} \right)^N e^{-\frac{100}{\beta} \cdot \sum_{i=1}^{N} \left|\frac{u'_i}{u} - 1\right|} \tag{4.5.2.14}$$

Posterior Distribution: Exponential Distribution Model

In the event of non-informative evidence regarding the unknown quantity, a uniform distribution $\pi_0(u)$ is assigned for the prior distribution:

$$\pi_0(u) = \frac{1}{u} e^{-\cdot\left|\frac{1}{u}\right|} \tag{4.5.2.15}$$

The resulting posterior distribution based on Equation 4.5.2.8 for the true value of the unknown quantity *u,* given the estimate of expert *"i"* yields:

$$\pi(u|u'_i) = \frac{\left(\frac{100}{u} \cdot \frac{1}{\beta} e^{-\frac{100}{\beta}\left|\frac{u'_i}{u}-1\right|}\right)\frac{1}{u} e^{-\cdot\left|\frac{1}{u}\right|}}{\int\left(\frac{100}{u} \cdot \frac{1}{\beta} e^{-\frac{100}{\beta}\left|\frac{u'_i-u}{u}\right|}\right)\frac{1}{u} e^{-\cdot\left|\frac{1}{u}\right|} du} = \frac{u^{-2} e^{-\frac{100}{\beta}\left|\frac{u'_i}{u}-1\right|} e^{-\cdot\left|\frac{1}{u}\right|}}{\int u^{-2} e^{-\frac{100}{\beta}\left|\frac{u'_i}{u}-1\right|} e^{-\cdot\left|\frac{1}{u}\right|} du} \tag{4.5.2.16}$$

and Equation 4.5.2.16 reduces to:

$$\pi(u|u'_i) = \frac{u^{-2} e^{-\frac{100}{\beta}\left|\frac{u'_i}{u}-1\right|}}{\frac{100}{u'_i\,\beta^{-1}}} = \frac{100u'_i}{\beta u^2} e^{-\frac{100}{\beta}\left|\frac{u'_i}{u}-1\right|} \tag{4.5.2.17}$$

The corresponding posterior distribution representing the decision maker's posterior state of knowledge about the unknown quantity u, given the estimates of multiple experts *1...N* is as follows:

$$\pi(u|u'_1 \ldots u'_N) = k^{-1}\left(\frac{100}{u\beta}\right)^N e^{-\frac{100}{\beta}\cdot\left|\frac{\sum_{i=1}^{N} u'_i}{u} - N\right|} \pi_o(u) \tag{4.5.2.18}$$

Let $\quad Y = \dfrac{1}{u} \qquad\qquad dY = -\dfrac{1}{u^2} du$

And the prior distribution of Y is non-informative, such that:

$$\pi\left(u\middle|u'_1...u'_N\right)=\frac{\left(Y\right)^N e^{-\frac{100}{\beta}\left|Y\sum_{i=1}^{N}u'_i\right|}}{-\int\left(Y\right)^{N-2} e^{-\frac{100}{\beta}\left|Y\sum_{i=1}^{N}u'_i\right|}dY} \tag{4.5.2.19}$$

$$\pi\left(u\middle|u'_1...u'_N\right)=\frac{\dfrac{\left(Y\right)^N e^{-\frac{1}{\beta}\left|Y\sum_{i=1}^{N}u'_i\right|}}{\Gamma\left(N-1\right)}}{\left(\dfrac{100\sum_{i=1}^{N}u'_i}{\beta}\right)^{N-1}} \tag{4.5.2.20}$$

$$\pi\left(u\middle|u'_1...u'_N\right)=\left(\frac{100\sum_{i=1}^{N}u'_i}{\beta}\right)^{N-1}\frac{exp\left[-\frac{1}{\beta}\cdot\left|\frac{\sum_{i=1}^{N}u'_i}{u}\right|\right]}{\left(u\right)^N\Gamma\left(N-1\right)} \tag{4.5.2.21}$$

The most likely value for *"u"*, denoted by $\hat{u}$ is derived by maximizing the posterior distribution (Equation 4.5.2.21) with respect to *"u"* and solving for *"u"*.

$$\frac{d\left[\pi\left(u\middle|u'_1...u'_N\right)\right]}{du}=0 \tag{4.5.2.22}$$

Likelihood Function:

Lognormal Distribution Model for a Single Expert

Empirical results from the meta-analysis of interdisciplinary expert judgment literature reveal that the absolute percentage error of expert estimates across disciplines are also consistent with a lognormal distribution. The standard lognormal probability density for a random variable *"x"*, with distribution parameters μ and σ is:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x - \mu)^2 / 2\sigma^2} \tag{4.5.2.23}$$

where:

μ is the mean of the logarithmic random variable x

σ is the standard deviation of the logarithmic random variable x

The normal probability distribution becomes lognormal when the log of the random variable x is normally distributed with mean μ and standard deviation σ.

The use of logarithmic likelihood functions to express or depict expert judgment is not a new concept. Mosleh (1981) used lognormal likelihood models to illustrate the multiplicative error relationship of a true value to an individual expert estimate, and multiple expert estimates. Mosleh (1981) assumed, the logarithmic random variable representing the estimate of the experts X_1 is the sum of the logarithmic true value $\ln x$ and error $\ln E_1$:

$$\ln X_1 = \ln x + \ln E_1$$

and $\ln X_1$, $\ln x$, and $\ln E_1$ are normally distributed.

From this relationship Mosleh et al (1981) developed the likelihood function of an expert's estimate given the true value of the unknown quantity for a single expert:

$$L(x_1|x) = \frac{1}{\sqrt{2\pi}\sigma_1 x_1} exp\left\{-\frac{1}{2}\left[\frac{\ln x_1 - (\ln x + \ln a)}{\sigma_1}\right]^2\right\} \tag{4.5.2.24}$$

where:

$\langle \ln X_1 \rangle = \ln x_1$, $\langle \ln x \rangle = \ln x$ and $\langle \ln E_1 \rangle = \ln a$

Mosleh et al (1981) further extended the logarithmic relationship to multiple experts. The derived joint likelihood function for multiple experts, with estimates vector $\underline{X}$ given the true value of the unknown quantity x below, is applicable to both independent and dependent experts:

$$L(\underline{x}|x) = \frac{1}{(2\pi)^{N/2}[det\,\Sigma]^{1/2}\prod_{i=1}^{N} x_i} \bullet \exp\left\{-\frac{1}{2}\left(ln\,\underline{X} - ln\langle\underline{x}\rangle\right)^T \Sigma^{-1}\left(ln\,\underline{x} - ln\langle\underline{x}\rangle\right)\right\} \quad (4.5.2.25)$$

where:

$ln\,\underline{X}$ is the logarithmic vector of expert estimates

$ln\,\underline{x}$ is the logarithmic vector of expected expert estimates

Σ is a matrix of covariance and directional dependency

$\left(ln\,\underline{X} - ln\langle\underline{x}\rangle\right)^T$ is the transpose of the difference of the logarithmic vectors $ln\,\underline{x}$ and $ln\,\underline{X}$

The findings of the meta-analysis support the logarithmic relationships developed by Mosleh et al. (1981), and have generated empirical values for model parameters. These empirical values contribute significantly to enhancing the logarithmic likelihood functions for expert judgment. Decision makers are now empowered with additional tools to assist in defining the relationship between expert estimates and the true values of unknown quantities.

Recall in Equation 4.5.2.2 that $u'_i = u(1 + 0.01E_i)$, where u'_i is the expert evidence, $"E_i"$ the percentage error, and *"u"* the true value of the unknown quantity.

As demonstrated in Chapter 3, the random variable denoting error, E, is a lognormal distribution:

$$f(E) = \frac{1}{E\sigma_E\sqrt{2\pi}} e^{-(\ln E - \ln \mu_E)^2 / 2\sigma_E^2} \tag{4.5.2.26}$$

This implies the expert's estimate u' is also a lognormal distribution. The corresponding likelihood function for expert evidence u', given that the true value of the unknown quantity is *"u"*, follows

$$L(u'|u) = \frac{1}{\sqrt{2\pi}\sigma_E \left|\frac{u'}{u} - 1\right| \cdot 100} \exp\left[-\frac{1}{2}\left(\frac{\ln\left(\left|\frac{u'}{u} - 1\right| \cdot 100\right) - \ln \mu_E}{\sigma_E}\right)^2\right] \tag{4.5.2.27}$$

The subsequent posterior distribution for the true value of the unknown quantity given the estimate u' from expert *"i"* is:

$$\pi(u|u') = k^{-1} \frac{1}{\sqrt{2\pi}\sigma_E \left|\frac{u'}{u} - 1\right| \cdot 100} \exp\left[-\frac{1}{2}\left(\frac{\ln\left(\left|\frac{u'}{u} - 1\right| \cdot 100\right) - \ln \mu_E}{\sigma_E}\right)^2\right] \pi_0(u) \tag{4.5.2.28}$$

Assuming the prior $\pi_0(u) = \frac{1}{u^2}$ for the unknown random variable *"u"*, the posterior distribution becomes

$$\pi(u|u') = \frac{\frac{1}{\sqrt{2\pi}\sigma_E\left(1-\frac{u'}{u}\right)} exp\left[-\frac{1}{2}\left(\frac{ln\left(\left|\frac{u'}{u}-1\right|\cdot 100\right) - ln\,\mu_E}{\sigma_E}\right)^2\right]\left(\frac{1}{u^2}\right)}{\int \frac{1}{\sqrt{2\pi}\sigma_E\left(1-\frac{u'}{u}\right)} exp\left[-\frac{1}{2}\left(\frac{ln\left(\left|\frac{u'}{u}-1\right|\cdot 100\right) - ln\,\mu_E}{\sigma_E}\right)^2\right]\left(\frac{1}{u^2}\right)du} \qquad (4.5.2.29)$$

Following the evaluation of the normalization constant, the posterior distribution in Equation 5.5.2.29 reduces to

$$\pi(u|u') = \frac{2u'}{\sqrt{2\pi}\sigma_E(u-u')u} exp\left[-\frac{1}{2}\left(\frac{ln\left|\frac{u'}{u}-1\right|\cdot 100 - ln\,\mu_E}{\sigma_E}\right)^2\right] \qquad (4.5.2.30)$$

The maximum likely value of *"u"* is derived from maximizing the log -posterior distribution such that:

$$\frac{\partial\, log[\pi(u|u')]}{\partial u} = 0$$

$$\frac{\partial}{\partial u}\left[-log\,u - log(u-u') - \frac{1}{2}\left(\frac{ln\left|\frac{u'}{u}-1\right|\cdot 100 - ln\,\mu_E}{\sigma_E}\right)^2\right] = 0 \qquad (4.5.2.31)$$

The maximum likely value for *"u"*:

$$\hat{u} = \frac{u'}{0.01 \times \mu_E e^{\sigma_E^2} + 1} \qquad (4.5.2.32)$$

In the ensuing steps, the lognormal likelihood function for a single expert is extended to multiple independent experts $1...N$. The lognormal likelihood function for multiple experts $1...N$ providing estimates $u'_1...u'_N$, given the true value of the unknown quantity is *"u"* is as follows:

$$L\left(u'_1...u'_N\left|u\right.\right)=$$

$$\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}\sigma_E\left|\frac{u'_i}{u}-1\right|\cdot 100}exp\left[-\frac{1}{2}\left(\frac{ln\left|\frac{u'_i}{u}-1\right|\cdot 100-ln\,\mu_E}{\sigma_E}\right)^2\right] \quad (4.5.2.33)$$

simplifying Equation 4.5.2.33 yields:

$$L\left(u'_1...u'_N\left|u\right.\right)=\left(\frac{1}{\sqrt{2\pi}\sigma_E}\right)^N\prod_{i=1}^{N}\left[\left|\frac{u'_i}{u}-1\right|\cdot 100\right]\times$$

$$exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{ln\left|\frac{u'_i}{u}-1\right|\cdot 100-ln\,\mu_E}{\sigma_E}\right)^2\right] \quad (4.5.2.34)$$

The posterior distribution of the unknown quantity *"u"* for multiple experts $1...N$, given their estimates $u'_1...u'_N$, developed below. The posterior distribution is formed from the likelihood function in Equation 4.5.2.34 and a non-informative prior $\pi_0\left(u\right)=\frac{1}{u^2}$, also used for the case of the single expert.

$$\pi\left(u\middle|u'_1...u'_N\right)=$$

$$\frac{\left(\frac{1}{100\sqrt{2\pi}\sigma_E}\right)^N\prod_{i=1}^{N}\left|\frac{u'_i}{u}-1\right|^{-1}exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{ln\left|\frac{u'_i}{u}-1\right|\cdot 100-ln\,\mu_E}{\sigma_E}\right)^2\right]\left(\frac{1}{u^2}\right)}{\int\left(\frac{1}{100\sqrt{2\pi}\sigma_E}\right)^N\prod_{i=1}^{N}\left|\frac{u'_i}{u}-1\right|^{-1}exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{ln\left|\frac{u'_i}{u}-1\right|\cdot 100-ln\,\mu_E}{\sigma_E}\right)^2\right]\left(\frac{1}{u^2}\right)du} \qquad (4.5.2.35)$$

Solving the normalization factor in Equation 4.5.2.35:

$$\int\left(\frac{1}{100\sqrt{2\pi}\sigma_E}\right)^N\prod_{i=1}^{N}\left(u'_i\right)^{-1}exp\left[-\frac{1}{2\sigma_E^2}\sum_{i=1}^{N}\left(Y_i-ln\,\mu_E\right)^2\right]dY_i \qquad (4.5.2.36)$$

where:

$$Y_i=ln\left(\left|\frac{u'_i}{u}-1\right|\cdot 100\right) \qquad \text{and} \qquad \frac{dY_i}{u'_i}=-\frac{1}{\left|\frac{u'_i}{u}-1\right|}\frac{1}{u^2}du$$

The normalization factor then reduces to:

$$\left(\frac{1}{100}\right)^N\left(\frac{1}{2}\right)^N\prod_{i=1}^{N}\frac{1}{u'_i} \qquad (4.5.2.37)$$

Substituting the normalization constant of Equation 4.5.2.37 into Equation 4.5.2.35 yields:

$$\pi\left(u\middle|u'_1...u'_N\right)=$$

$$\left(\frac{2}{\sqrt{2\pi}\sigma_E}\right)^N\prod_{i=1}^{N}\left|\frac{u'_i}{u'_i-u}\right|exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{ln\left|\frac{u'_i}{u}-1\right|\cdot 100-ln\,\mu_E}{\sigma_E}\right)^2\right]\left(\frac{1}{u}\right) \qquad (4.5.2.38)$$

The most likely value for *"u"*, denoted by $\hat{u}$ is derived by maximizing the log of the posterior distribution (Equation 4.5.2.38) with respect to *"u"* and solving for *"u"*.

$$\frac{\partial}{\partial u}\left[log\left[\pi\left(u \middle| u'_1 ... u'_N\right)\right]\right] = 0 \qquad (4.5.2.39)$$

This implies:

$$\frac{\partial}{\partial u}\left[\sum_{i=1}^{N} log\left(\frac{u}{u'_i - u}\right) - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{ln\left(\left|\frac{u'_i}{u} - 1\right| \cdot 100\right) - ln\,\mu_E}{\sigma_E}\right)^2 - \sum_{i=1}^{N} log\,u\right] = 0 \qquad (4.5.2.40)$$

The following is the resulting most likely value for the unknown quantity *u*, given expert estimates $u'_1 ... u'_N$:

$$\hat{u} = \frac{\sum_{i=1}^{N} u'_i}{N + 0.01 \times \mu_E e^{\sigma_E^2}} \qquad (54.5.2.41)$$

Equation 4.5.2.32 is a special case of equation 4.5.2.41

4.7 SUMMARY

An expert's estimate is associated with a certain measure of error, by means of its empirical relationship to the true value of the relevant quantity. Error is defined in this text as the ratio of the difference in a true value and an estimate over the true value of a particular quantity. The meta-analysis presented in Chapter 3 reveals that two distributions, the exponential and lognormal, are good fits for the error data set used in this analysis. As a result, Bayesian relationships for the true value of unknown quantities were derived from the two distributions. This entailed the development of the likelihood functions of the expert evidence given the true value of the quantities, along with the posterior distributions for the quantities of interest. In addition, the equations for the most likely value of the unknown quantities of interests were derived.

Optimistic validation results of the equations for the determination of the true value of unknown quantities would serve potentially significant to decision makers and stakeholders. Given the estimate of an expert, these equations can potentially inform the user of the likely values of the unknown quantity. These equations are dependent on the parameters of the lognormal distributions. The author is aware of the uncertainty surrounding these parameters, and recommends the use of the calibrated posterior distribution for the unknown quantity. The calibrated posterior distribution is able to better inform the decision maker of true value of the unknown quantity. Also, the unknown quantity can be determined as a point estimate or a distributed value. The mean and the variance of the posterior distribution can provide the parameters of the distribution for the unknown quantity.

CHAPTER 5

EVALUATION OF EXPERT PERFORMANCE WITH AND WITHOUT THE CONTRIBUTION OF EXPERT ATTRIBUTES

5.1 ABSTRACT

The correlation of attributes to performance has been explicitly expressed in various case studies. Studies have concluded or recommended the selection of experts with various attributes to yield improved accuracy. Other studies have conducted research to empirically show the correlation between attributes and performance. A few of these studies showed incremental increase in performance. However, these studies failed to present an interdisciplinary framework to predict or forecast expert performance or accuracy. In this chapter nonlinear multiple regression equations were developed to forecast expert performance, based on expert attributes

5.1 INTRODUCTION

Expert judgments are an essential part of the decision-making process. As a result, decision makers and stakeholders across all disciplines consistently ask the age-old question, "How accurate are the experts?" This question has gone relatively unanswered, with the exception of cases wherein historical performance data on individual experts are available. Resolution to this question is hindered by several factors, including inadequate descriptions of expert attributes in the published validated expert judgment case studies, as well as non-standard measurement of expertise. Proper assessment of expert performance or accuracy was achievable by a two-step process. First, there must be a good understanding of the critical layers impeding the solution to the question. Second, the creation of a framework to remove these layers is essential. This chapter presents the development of a novel tool which enables decision makers to predict expert performance, based on the attributes or qualifications of experts.

Attributes are the characteristics, traits, and peculiarities relating to an individual. Distinction between experts and novices are conducted based on specific attributes. The perception of the quality attributes that make a person an expert is subject and unique to individuals. Weiss and Shanteau (2003) identified experts by attributes such as self-proclamation or peer nomination as well as by experience, titles, and degrees. Additional attributes include membership in professional organizations, number of expert participants and characteristics of tasks such as frequency of occurrence and difficulty (Stewart et al. 1997).

5.2 METHODOLOGY

The development of a mathematical framework to assess attribute-dependent expert performance was created from a meta-analysis of literature. The most commonly recommended attributes of experts were identified from literature. A search of literature for validated expert judgment case studies resulted in eighty data sets. Each case study was evaluated for the following attributes: peer nomination, certification or specialized training in expertise, publications in expertise or field, as well as membership in professional organizations, organization specialization, institution type, average level of formal education, and average years of experience. Descriptions of experts, as well as elicited and validated judgments were evaluated to determine the relationship of predictors or attributes, and performance.

The search for the accuracy of expert opinion or expert judgment began with a general survey of past and current literature, internet publications, refereed and non-refereed sources. The wide literature search included in part the following databases: WorldCat, Agricola, DOE's Information Bridge, Civil Engineering (CE) database, Energy Citations database, Waste Management research abstracts, PubMed and Medline. The most insightful abstracts were found in PubMed, Medline, and WorldCat. In addition, a worldwide exploration of the Dissertation Abstracts database was performed to identify any similar or exact work across all disciplines. One hundred ninety one dissertation abstracts were deemed relevant and selected for review.

The number of articles reviewed and selected is in Appendix F, and descriptions of all databases are in Appendix G. The search generated over 3000 abstracts, the most relevant (over 1000) abstracts were organized into five categories and selected for review. The categories are forecasting, medical sciences, litigation, plant science, and expert systems.

Selections or rejections of articles were performed in two phases. The first phase occurred at the database and other search engines level, where available abstracts, book summaries, and sources without summaries or abstracts, were examined for relevance. Those that were seemingly or at worst partially relevant were selected. During the second phase, articles and books identified in Phase one were obtained from various libraries and internet sites and the full contents of the sources examined for information/ parameters/ data listed in Appendix F. Sources with the information in the form as printed in the Appendix A are separated. Next, sources with relevant information to be later converted into standards required in Appendix A are selected. The remaining sources were searched for supplemental information or discarded.

The data gathered were then organized into a SPSS (Statistical Package of the Social Sciences) compatible matrix for correlation studies. Each study was evaluated in part for quality attributes such as education, organizational membership, and nomination by peers, and type of institution they are associated with. SPSS 10.0 was used in the initial analysis of the data. Following the analysis additional literature searches were conducted to clarify and cover data gaps. The authors of sources lacking relevant information were contacted for additional data. Several rounds of analysis and additional research were subsequently performed until optimal results were achieved.

The resulting searches did not generate any previous work closely matching the research performed in this study. In addition, the need for further research into the correlation of expert quality attributes and their subsequent accuracy was reinforced. The search of literature yielded much information of various expert judgment accuracy studies. Several studies quantified the accuracy of expert judgment for specific diseases and various fields. Despite relatively few attempts made to broadly and quantitatively evaluate the relationship between quality attributes and error, there exist several studies wherein expert opinions and expert systems were validated.

The selected expert opinion studies were in several forms. Some studies were complete with detailed descriptions of the experts quality attributes along with the accuracy of their judgments. Others were partially or vaguely descriptive with such language as simply “expert”, “experts with many years of experience”, etc. The analyses of these two primary types of data sets were analyzed separately and will be discussed further in this section.

The likelihood functions generated by this meta-analysis were the result of expert judgment for categorical quantities. Categorical quantities have a finite number of possible values within a range, and may be further subdivided into ordinal or unordered quantities. Ordinal quantities contain a hierarchy, such as low, medium, and high. Examples of categorical quantities include, condition present or absent, correct or incorrect diagnosis, and selection of specific condition. The studies reported percentage accuracies for experts on specific issues and sub-disciplines. These results were fitted to distributions using the Best Fit software. Other relationships were generated by nonlinear multiple regression analyses. Solutions to the likelihood functions and posterior distributions were developed for both single and multiple experts. Decision makers are keenly interested in both types of functions.

5.3 PROBABILITY DENSITY FUNCTIONS FOR PERCENT ACCURACY OF EXPERTS:

As previously stated in Chapter 3, the beta and the right-truncated normal distributions illustrate the probability density of overall accuracy for independent experts. The uncertain random variable *"p"* denotes the probability of expert(s) making correct prediction(s), and is governed by the parameters of the distributions. Consider an occasion wherein a decision maker possesses limited information about an expert, but desires to know the performance of that expert. In this section, the calibrated posterior distributions for expert performance accuracy are presented.

Beta Distribution

The beta distribution is a two parameter (*a* and *b*) probability density function. The parameters *a* and *b* describe the shape of the distribution. Given an expert's claim that he or she is correct, the probability *"p"* of expert(s) correctly predicting all states of nature is represented by the beta distribution below:

$$\pi(p|a,b) = k^{-1} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} (p)^{a-1} (1-p)^{b-1} \tag{5.3.1}$$

Controlled by the distribution of *a* and *b*, the optimal solution for *"p"* is indicated by:

$$\pi(p|\hat{a},\hat{b}) = \frac{\Gamma(\hat{a}+\hat{b})}{\Gamma(\hat{a})\Gamma(\hat{b})} (p)^{\hat{a}-1} (1-p)^{\hat{b}-1} \tag{5.3.2}$$

or

$$\pi(p|\hat{a},\hat{b}) = \iint_{a,b} \pi(p|a,b)\pi(a,b|evidence)dadb \tag{5.3.3}$$

Derived from the evidence of expert percentage accuracies in Chapter 3, the resulting distribution for both model parameters *a* and *b* are developed in the ensuing steps, such that:

$$\pi(a,b) = \frac{L(p_1....p_N|a,b)\pi_0(a,b)}{\iint_{a,b} L(p_1....p_N|a,b)\pi_0(a,b)dadb} \tag{5.3.4}$$

The likelihood of function for the observation of the accuracy p_i of expert *"i"* given, the true value of model parameters *a* and *b* is as follows:

$$L(p_i|a,b) = (B(a,b))((p_i)^{a-1}(1-p_i)^{b-1}) \tag{5.3.5}$$

Likewise, for multiple experts *1…N,* the equation below depicts the likelihood of observing expert accuracies $p_1 \ldots p_N$ for experts *1…N*, given that the true value for the overall interdisciplinary expert accuracy:

$$L(p_1,\ldots p_N|a,b) = \prod_{i=1}^{N} L(p_i|a,b) = (B(a,b))^N \prod_{i=1}^{N} \left((p_i)^{a-1}(1-p_i)^{b-1}\right) \quad (5.3.6)$$

The maximum likelihood estimators for *a* and *b* are derived by maximizing Equation 5.3.6 and solving for a and b, respectively:

$$\frac{\partial}{\partial a}\left[L(p_1,\ldots p_N|a,b)\right] = 0$$
$$\frac{\partial}{\partial b}\left[L(p_1,\ldots p_N|a,b)\right] = 0 \quad (5.3.7)$$

Assuming a uniform prior distribution, the posterior distribution for model parameters *a* and *b* in Equation 5.3.4 becomes:

$$\pi(a,b) = \frac{\prod_{i=1}^{N} L(p_i|a,b)}{\iint_{a,b} \prod_{i=1}^{N} L(p_i|a,b)\,dadb} \quad (5.3.8)$$

or

$$\pi(a,b) = \frac{(B(a,b))^N \prod_{i=1}^{N} \left((p_i)^{a-1}(1-p_i)^{b-1}\right)}{\iint_{a,b} (B(a,b))^N \prod_{i=1}^{N} \left((p_i)^{a-1}(1-p_i)^{b-1}\right) dadb} \quad (5.3.9)$$

The posterior distribution of Equation 5.3.9 represents the uncertainty in parameters *a* and *b* of the beta distribution representing expert performance (Equation 5.3.1).

Consequently, the calibration equation for *"p",* the probability of expert(s) making correct prediction(s) can be obtained from integrating the product of Equations 5.3.1 and 5.3.9 over *a* and *b*:

$$\pi\left(p\middle|\hat{a},\hat{b}\right)= \iint_{a,b}\left\{\left[\frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(p)^{a-1}(1-p)^{b-1}}{\int\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(p)^{a-1}(1-p)^{b-1}dp}\right]\bullet \left[\frac{(B(a,b))^{N}\prod_{i=1}^{N}\left((p_i)^{a-1}(1-p_i)^{b-1}\right)}{\iint_{a,b}(B(a,b))^{N}\prod_{i=1}^{N}\left((p_i)^{a-1}(1-p_i)^{b-1}\right)dadb}\right]\right\}dadb \qquad (5.3.10)$$

Solutions to Equation 5.3.7 through 5.3.10 can be achieved numerically. In addition, the most likely value for *"p"* is attained by maximizing Equation 5.3.10 and solving for *"p"*.

<u>Truncated Normal Distribution</u>

As illustrated in Figure 4 of Chapter 3, the distribution denoting expert performance or percent accuracy is also consistent with the shape of a normal distribution, truncated on the right. A normally-distributed random variable *"x"* with a probability density function $f(x)$ is specified as:

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{for } -\infty \le x \le \infty \qquad (5.3.11)$$

if $$z=\frac{x-\mu}{\sigma}, \qquad (5.3.12)$$

then, the standard normal distribution is denoted by:

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2} \tag{5.3.13}$$

In special cases wherein the normal distribution is truncated on the right or left side, the terms as x_R or x_L designate those points, respectively. The standard normal terms denoting the truncation points are:

$$k_R=\frac{x_R-\mu}{\sigma} \qquad \text{for the right} \tag{5.3.14}$$

$$k_L=\frac{x_L-\mu}{\sigma} \qquad \text{for the left} \tag{5.3.15}$$

The standard right truncated normal distribution is therefore expressed as:

$$f_{TR}(t)=\begin{cases}\dfrac{f(t+k_R)}{\int_0^{k_R} f(z)dz} & \text{for } t\le 0\end{cases} \tag{5.3.16}$$

where $t=z-k_R$

and Equation 5.3.16 reduces to:

$$f_{TR}(t)=\begin{cases}\dfrac{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\int_0^{k_R}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z)^2}dz}\end{cases} \tag{5.3.17}$$

The form of the normalization factor in Equation 5.3.17 is consistent with the *continued fraction identity* below:

$$\int_0^y e^{-u^2} du = \frac{\sqrt{\pi}}{2} - \cfrac{\frac{1}{2}e^{-y^2}}{y + \cfrac{1}{2y + \cfrac{2}{y + \cfrac{3}{2y + \cfrac{4}{y + \cfrac{5}{2y + \cfrac{6}{y + \dots}}}}}}} \tag{5.3.18}$$

Applying the continued fraction identity to Equation 5.3.16, and approximating in Equation 5.3.17, the right truncated normal distribution reduces to:

$$f_{TR}(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2} \cdot \left(\frac{\sqrt{\pi}}{2} - \frac{exp\left(-(k_R)^2\right)}{2k_R} \right)} \tag{5.3.19}$$

or substituting Equation 5.3.14 into Equation 5.3.19 yields

$$f_{TR}(x) = \frac{2(x_R - \mu)e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2} \cdot \left(\sqrt{\pi}(x_R - \mu) - \sigma\, exp\left(-\left(\frac{x_R - \mu}{\sigma}\right)^2 \right) \right)} \tag{5.3.20}$$

Applying Equation 5.3.20 to the distribution of *"p"* obtained in Chapter 3 result in the following posterior distribution given μ_P and σ_P:

$$\pi(p|\mu_P, \sigma_P) = \frac{(100-\mu_P)^N \left(\sqrt{\pi}(100-\mu_P) - \sigma_P e^{-\left(\frac{100-\mu_P}{\sigma_P}\right)^2} \right)^{-N} e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{p_i - \mu_P}{\sigma_P}\right)^2}}{\int (100-\mu_P)^N \left(\sqrt{\pi}(100-\mu_P) - \sigma_P e^{-\left(\frac{100-\mu_P}{\sigma_P}\right)^2} \right)^{-N} e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{p_i - \mu_P}{\sigma_P}\right)^2} dp} \tag{5.3.21}$$

Single expert *"i":*

The resulting normal likelihood function truncated at $p = 100\%$, reflects the observance of percent accuracy $"p"$ for expert *"i"*, given the true value for model parameters μ_P and σ_P, and is denoted:

$$L(p|\mu_P,\sigma_P) = \frac{2(100-\mu_P)e^{-\frac{1}{2}\left(\frac{p-\mu_P}{\sigma_P}\right)^2}}{\sqrt{2}\cdot\left(\sqrt{\pi}(100-\mu_P)-\sigma_P\, exp\left(-\left(\frac{100-\mu_P}{\sigma_P}\right)^2\right)\right)} \tag{5.3.22}$$

Multiple experts *1…N:*

The likelihood of observing expert accuracies $p_1 \ldots p_N$ for experts *1…N*, given that the true value for the model parameters is μ_P and σ_P:

$$L(p_1 \ldots p_N|\mu_P,\sigma_P) =$$

$$\prod_{i=1}^{N}\frac{2(100-\mu_P)e^{-\frac{1}{2}\left(\frac{p_i-\mu_P}{\sigma_P}\right)^2}}{\sqrt{2}\cdot\left(\sqrt{\pi}(100-\mu_P)-\sigma_P\, exp\left(-\left(\frac{100-\mu_P}{\sigma_P}\right)^2\right)\right)} \tag{5.3.23}$$

Equations 5.3.22 and 5.3.23 reflect a single point estimate for μ_P and σ_P, for the truncated normal function. However, for any given expert or experts, there exists some uncertainty in the true value of *"p"*. The associated uncertainties are express in the distribution of μ_P and σ_P. The following posterior distribution for μ_E and σ_E, expresses the associated uncertainty in *"p"*:

$$\pi(\mu_P,\sigma_P|p_1 \ldots p_N) = \frac{L(p_1 \ldots p_N|\mu_P,\sigma_P)\pi_0(\mu_P,\sigma_P)}{\iint_{\mu_P,\sigma_P} L(p_1 \ldots p_N|\mu_P,\sigma_P)\pi_0(\mu_P,\sigma_P)d\mu_P d\sigma_P} \tag{5.3.24}$$

Assuming a uniform prior for the distribution of μ_E and σ_E, the posterior distribution yields:

$$\pi\left(\sigma_P,\mu_P \middle| p_1....p_N\right)=$$

$$k^{-1}\frac{\left(\sqrt{2}\left(100-\mu_P\right)\right)^N e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{p_i-\mu_P}{\sigma_P}\right)^2}}{\left(\sqrt{\pi}\left(100-\mu_P\right)-\sigma_P\, exp\left(-\left(\frac{100-\mu_P}{\sigma_P}\right)^2\right)\right)^N} \qquad (5.3.25)$$

where:

$$k=\iint_{\sigma_P,\mu_P}\frac{\left(\sqrt{2}\left(100-\mu_P\right)\right)^N e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{p_i-\mu_P}{\sigma_P}\right)^2}}{\left(\sqrt{\pi}\left(100-\mu_P\right)-\sigma_P\, exp\left(-\left(\frac{100-\mu_P}{\sigma_P}\right)^2\right)\right)^N}\sigma_P,\mu_P \qquad (5.3.26)$$

The calibrated posterior distribution, below, best reflects the true value of expert performance *"p"*, and its associated uncertainties.

$$\pi\left(p\middle|\hat{\mu}_P,\hat{\sigma}_P\right)=\iint_{\mu_P,\sigma_P}\pi\left(p\middle|\mu_P,\sigma_P\right)\pi\left(\mu_P,\sigma_P\middle|p_1...p_N\right)d\mu_P d\sigma_P \qquad (5.3.27)$$

The solution to the calibration equation and the most likely values of the distribution parameters can be evaluated numerically.

5.4 REGRESSION MODELING OF PERCENT ACCURACY OF EXPERTS: (ATTRIBUTES DEPENDENCE)

The theoretical association of expert attributes and performance is well-established. However, empirical models to forecast expert performance based on attributes are limited to very specific cases with historical expert performance of each individual. This section presents an interdisciplinary regression relationship between experts' attributes and accuracy, based on the data acquired and analyzed in Chapter 3. The attributes used to formulate the regression relationship are the most commonly recommended in literature.

A regression equation predicts values of a dependent variable from at least one independent variable. The standard linear regression equation is expressed below:

$$Y = \alpha + \beta X + \varepsilon \tag{5.4.1}$$

where:

Y denotes the dependent, or response continuous variable (The response variable is generally random, but may not be)

X denotes the independent, or explanatory variable(s), or covariate

α is the intercept

β is the slope or regression coefficient

ε is the error term

Or the regression equation could be denoted:

$$Y = f(X) + \varepsilon = f(X_1 \ldots X_N ; a_1 \ldots a_n) + \varepsilon \tag{5.4.2}$$

where:

$f(X_1 \ldots X_N)$ is the regression function containing N number of row vectors of explanatory variables

The mathematical relationship formed in this research, between attributes and accuracy is consistent with nonlinear multiple regression. Nonlinear regression differs from linear in that the relationship between a response variable and one or more explanatory variables is non-linear, and it permits any continuous or discontinuous model. In addition, this type of regression facilitates multiple states of the explanatory variables. The regression relationship of attributes to expert performance is denoted in the equation below:

$$p = \beta\, ln\left|X\right| + T \qquad (5.4.3)$$

where:

p denotes expert accuracies $p_1,...p_N$, a set of observed dependent variables

X denotes a vector of explanatory variables or attributes

β denotes the population slop

T is either fixed or random. As a random variable, it denotes a distribution of the possible task-related attributes contributing to expert accuracy.

Equation 5.4.1 mathematically depicts the accuracy of experts $''p''$ to a logarithmic relationship of the set of expert attributes $''X''$ and task related attributes T of experts.

In this section two vectors for $''X''$ corresponding to the curve fittings in Chapter 3 are presented. The first relationship denotes vector $''X''$ as a summation of attributes. The second expresses vector $''X''$ as a set of explanatory variables $\left(x_{i1},...x_{iN}\right)$ and corresponding correlation coefficients $\left(c_1,...c_N\right)$.or individual attributes. The correlation coefficients express the strength of the relationship of each attribute to accuracy, the values of which were presented in Chapter 3.

For the first relationship, all attributes are assumed to be fully correlated to expert performance, wherein $c_{ij} = 1$ such that:

$$X_i = \sum_{j=1}^{N} c_{ij} x_{ij} = \sum_{j=1}^{N} x_{ij} \qquad (5.4.2)$$

where:

x_{ij}	is the j^{th} attribute in set $''X_i''$ for expert $''i''$
$x_{ij} = m$	is the m^{th} state (i.e. attribute present) of the j^{th} attribute of expert $''i''$ in set $''X_i''$
$x_{ij} = n$	is the n^{th} state (i.e. attribute absent) of the j^{th} attribute of expert $''i''$ in set $''X_i''$

The regression equation $p_i = \beta\, ln\left|\sum_{j=1}^{N} x_{ij}\right| + T$ is bounded by the following condition:

$$0 < X_i \leq e^{(p_i - T)/\beta} \qquad \text{and} \qquad P_i = 100 \tag{5.4.3}$$

In the second relationship:

$$X_i^C = \sum_{j=1}^{N} x_{ij} c_{ij} \tag{5.4.4}$$

where:

c_{ij} is the coefficient describing the strength of the correlation between attribute x_{ij} and expert performance or accuracy $''p_i''$

The meta-analysis data was evaluated for relationships of accuracy and log of accuracy to attributes. Given these two conditions the regression equations were bounded by the following conditions.

Given $p_i = \beta\, ln\left|\sum_{j=1}^{N} x_{ij} c_{ij}\right| + T$

$$0 < X_i^C \leq e^{(ln\, p_i - T)/\beta} \qquad \text{and} \qquad p_i = 100 \tag{5.4.5}$$

Given $ln\, p_i = \beta\, ln\left|\sum_{j=1}^{N} x_{ij} c_{ij}\right| + T$

$$0 < X_i^C \leq e^{(P_i - T)/\beta} \qquad \text{and} \qquad P_i = 100 \tag{5.4.6}$$

Empirical values of $"\beta"$ and $"T"$ derived in Chapter 3 are based on the data set attained from literature. The data set is an appropriate representation of published validation studies. However, there is some uncertainty about the true values of $"\beta"$ and $"T"$, resulting in an array of possible fits for $"p"$.

$$\pi(p|X) = f(\beta, T) \tag{5.4.7}$$

This implies for any given value of $"X"$, there exists a distribution of $"p"$ values (see Figure 9).

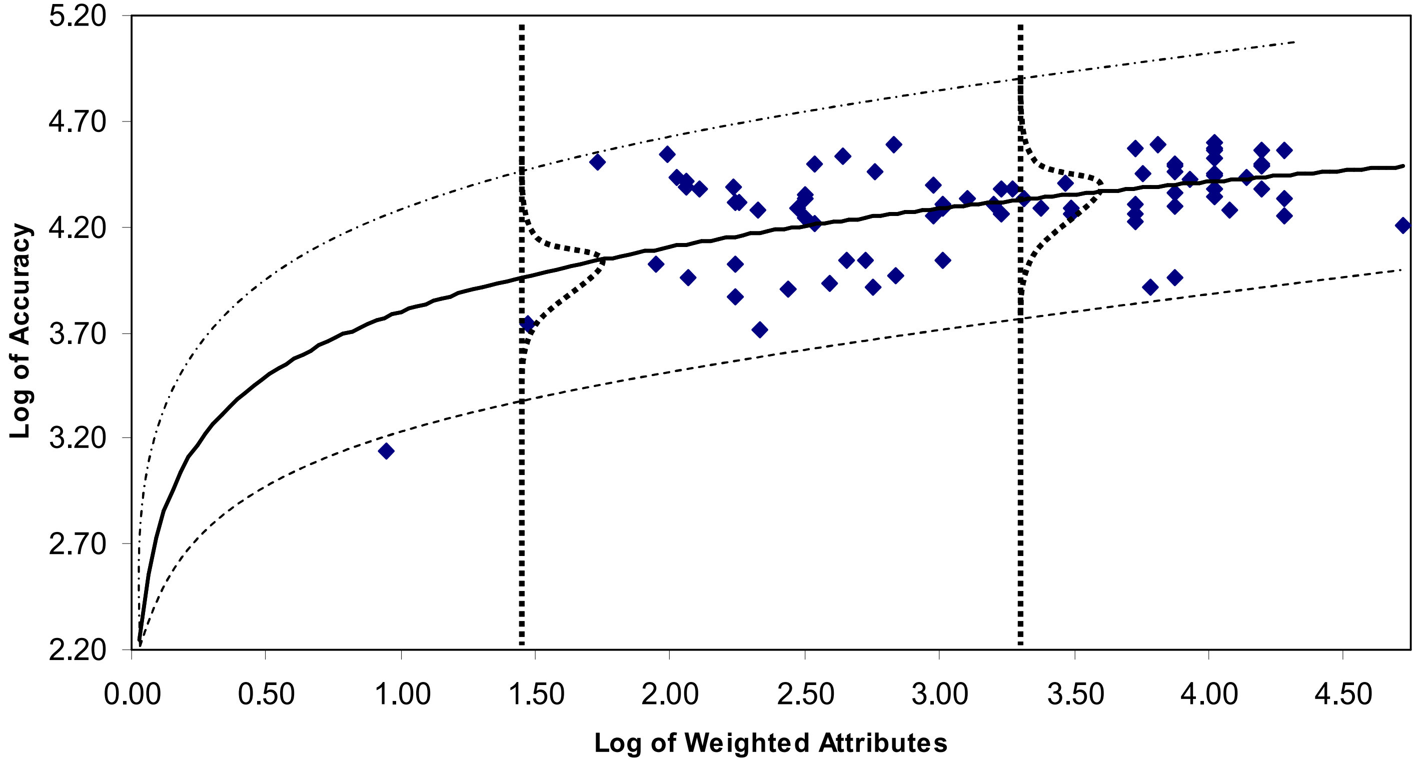

Figure 9: Logarithmic Fits of Accuracy Illustrating Uncertainty in $''\beta''$ and $''T''$

5.7 SUMMARY

The results of the meta-analysis revealed that expert performance can be expressed in the form of a beta or right truncated normal distribution. However, the true value for expert accuracy or performance is uncertain, and is governed by the model parameters. As a result, calibrated posterior distributions for expert accuracy were evaluated and are recommended. Given an experts' confidence that he or she is correct regarding a particular judgment, these equations may be applied to the prediction of expert performance.

Another result of the meta-analysis is the development of non-linear multiple regression relationships between expert attributes and performance. The relationships indicate that performance is a function of the log of an array of attributes. Within this array, the product of each individual attribute and its corresponding correlation coefficient are summed. This implies, given the qualifications or attributes of an expert, performance can be predicted.

The regression equations currently allow for the evaluation of experts for the following attributes: peer nominations, certification or specialized training in expertise, publications expertise or field, membership in professional organizations, and organization specialization in expertise, as well as institution type, average level of formal education, event frequency, and average years of experience. However, the equations also permit the inclusion of additional attributes; limitations on the total number of attributes are governed by a maximum value for the array of attributes, *"X"* in Equation 5.4.6.

There exists a distribution of possible values for percent accuracy *"p"*, or the probability of an expert correctly providing judgment(s). The uncertainty in *"p"*, is governed by the parameters $"\beta"$ and $"T"$ (see Equation 5.). The validity of the nonlinear multiple regression equations will be evaluated through case studies in Chapter 6.

CHAPTER 6

MODEL VALIDATION: CASE STUDIES

6.1 INTRODUCTION

This Chapter focuses on the validation of the equations for estimating the true values of an unknown quantity, along with the performance of experts based on their attributes. The validations were performed on two primary and two secondary case studies. The two primary studies were administered by the author in collaboration with the Department of Meteorology and Atmospheric Sciences at the University of Maryland-College Park, and the other with the American Dietetic Association. The secondary case studies were attained from Katy Walker, an independent consultant to the US-Environmental Protection Agency, and Ali Mosleh from the University of Maryland.-College Park. These case studies covered the varied disciplines of Meteorology, Environmental Science, Human Nutrition, and Engineering. Validation of the equations across disciplines would support their legitimacy.

6.2 VALIDATION OF LIKELIHOOD FUNCTIONS FROM EXPERT JUDGMENT CASE STUDIES

Given an estimate or a number of estimates regarding unknown quantities, Equation 4.5.2.32 has been developed to evaluate the true value of these quantities. In this section, data obtained from the Benzene Concentration, Engineering, Human Nutrition, and Forecasting case studies are used to evaluate the validity of Equation 4.5.2.43. Initial validation exercises of the exponential distribution being a good predictor of unknown quantities, given the estimates of experts, revealed very poor fits for the data sets. As a result, validation of equations derived from the exponential fits was not included in this Chapter.

6.2.1 Benzene Case Study*

A Secondary Human Exposure Assessment Survey (NHEXAS) was conducted to obtain exposure assessment experts judgments about uncertainty in residential ambient, residential indoor, and personal air benzene concentrations in U.S. EPA's Region V, experienced by the nonsmoking, nonoccupationally exposed population. The judgments of seven experts were elicited regarding the means and 90th percentiles of each of the benzene concentration. These experts were selected by a peer nomination process. In addition, individually elicited judgments were gathered from the experts during a 2-day workshop. "Specifically, each expert was asked to characterize, in probabilistic form, the arithmetic means and the 90th percentiles of these distributions."

* Walker, K., Catalano, P., Hammitt, J., and Evans, J (2003). Use of expert judgment in exposure assessment: Part 2. Calibration of expert judgments about personal exposures to benzene. Journal of Exposure Analysis and Environmental Epidemiology 13: 1-16.

Table 9. Expert Estimates of Ambient Benzene Concentrations (ug/m^3) versus Lognormal Model Projections

		Lognormal Model Adjusted Expert Estimate	
Expert	**Expert Estimate**	**Lower Bound**	**Upper Bound**
A	3.9	2.2	15.7
B	3.2	1.8	12.7
C	4.6	2.6	18.7
D	7.8	4.5	31.5
E	5.8	3.3	23.5
F	3.2	1.9	13.1
G	3.7	2.1	15
Mean	**4.6**	**2.6**	**18.6**

The true value for the Ambient Benzene Concentration is: 3.6 ug/m^3

Table 10. Expert Estimates of Indoor Benzene Concentrations ($\mu g/m^3$) versus Lognormal Model Projections

		Lognormal Model Adjusted Expert Estimate	
Expert	**Expert Estimate**	**Lower Bound**	**Upper Bound**
A	5.5	3.15	22.2
B	6.2	3.5	25.0
C	6.5	3.7	26.0
D	16.2	9.3	65.3
E	15.6	8.9	62.7
F	11.2	6.4	45.2
G	6	3.4	24.2
Mean	**9.6**	**5.5**	**38.6**

The true value for the Indoor Benzene Concentration is: 7.2 ug/m^3

Table 11. Expert Estimates of Personal Benzene Concentrations (ug/m^3) versus Lognormal Model Projections

		Lognormal Model Adjusted Expert Estimate	
Expert	**Expert Estimate**	**Lower Bound**	**Upper Bound**
A	13.9	7.9	55.8
B	7	4.0	28.1
C	8.6	4.9	34.5
D	11.2	6.4	44.9
E	21.7	12.4	87.5
F	12.1	6.9	48.6
G	7.9	4.5	31.8
Mean	**11.8**	**6.7**	**47.3**

The true value for the Personal Benzene Concentration is: 7.5 ug/m^3

Tables 9 through 11 contain the actual estimates of experts, the model projections the true value of the quantities. Mean expert estimate for the ambient, indoor and personal concentrations, were 4.6, 9.6 and 11.8, respectively; the true values were 3.6, 7.2 and 7.5, respectively. The lognormal model lower bound adjustments of expert estimates were 2.6, 5.5 and 6.7, respectively. Similarly, the upper bound estimates were 18.6, 38.6 and 47.3, respectively. The results indicate, the lower bound values were closer to the true value of the quantity.

6.2.2 Weather Precipitation Case Study

A weather precipitation case study among expert meteorologists at the University of Maryland- College Park was performed. The objectives of the study were three-fold: (1) to predict the APE of experts given their estimates and (2) to determined the effect of expertise on expert performance. The third objective of the study will be discussed in the next chapter.

The case study involved four experts who were asked to make 48 hour precipitation forecasts projections. In the field of meteorology, a 48hour forecast of precipitation is considered moderately difficult, and requires specialized skills. The forecast were conducted on three different days for the following cities: Orlando, Seattle, San Francisco, New Orleans and Detroit. The complete record of the survey instrument is documented in Appendix D.

Table 12a. Experts and Model Adjusted Estimates of the 48hrs Precipitation Forecast at the International Airports in Orlando, Seattle, San Francisco, New Orleans, Detroit

CITIES	Expert Forecast					Adjusted Expert Forecast: Lognormal Model Lower Bound				
	1	2	3	4	Mean	1	2	3	4	Mean
Seattle-D1	0.9	0.12	0.47	0.77	**0.565**	0.51	0.069	0.27	0.44	**0.32**
San Francisco-D1	0	0	0	0.01		0	0	0	0.006	
Seattle-D2	0	0.35	0.39	0.68	**0.4733**	0	0.20	0.22	0.39	**0.27**
Seattle-D3	0.5	----	0.79	0.66	**0.65**	0.29	----	0.45	0.38	**0.37**
Detroit-D3	0.1	----	0.04	0.2	**0.1133**	0.057	----	0.023	0.11	**0.065**

Table 12b. True Precipitation at the International Airports in Orlando, Seattle, San Francisco, New Orleans, and Detroit

	Seattle-D1	San Francisco-D1	Seattle-D2	Seattle-D3	Detroit-D3
True Precipitation	0.43	0.01	0.07	0.87	0.08

Table 13a. Experts and Model Adjusted Estimates of the 48hrs Precipitation Forecast at the International Airports in Orlando, Seattle, San Francisco, New Orleans, Detroit

CITIES	Expert Forecast					Adjusted Expert Forecast: Lognormal Model Upper Bound				
	1	2	3	4	Mean	1	2	3	4	Mean
Seattle-D1	0.9	0.12	0.47	0.77	**0.565**	3.6	0.48	1.9	3.1	**2.27**
San Francisco-D1	0	0	0	0.01		0	0	0	0.04	
Seattle-D2	0	0.35	0.39	0.68	**0.4733**	0	1.4	1.6	2.74	**1.43**
Seattle-D3	0.5	----	0.79	0.66	**0.65**	2.01	----	3.2	2.66	**2.62**
Detroit-D3	0.1	----	0.04	0.2	**0.1133**	0.40	----	0.16	0.81	**0.46**

Table 13b. True Precipitation at the International Airports in Orlando, Seattle, San Francisco, New Orleans, and Detroit

	Seattle-D1	San Francisco-D1	Seattle-D2	Seattle-D3	Detroit-D3
True Precipitation	0.43	0.01	0.07	0.87	0.08

Results in Table 9 shows the precipitation estimates of the four experts for precipitation in the cities of San Francisco, Detroit and Seattle. Of the five precipitation forecasts, three of the model's lower bound means precipitation estimates were better than that of the experts.

6.2.3 Component Maintenance Case Study

The Component Maintenance secondary case study obtained from Ali Mosleh from the University of Maryland, looked at mechanical equipment, electrical and electronic component failure rates. Table 10 displays results of the mean expert judgment, historic data, and the lognormal based estimations of component maintenance times. The lognormal model estimates are adjustments of expert judgments. Table 10 shows that 9 of the 12 model adjustments were closer to the true value of the quantity than the original expert estimates.

Table 13. Comparison of Data and Expert Opinion on the Mean Distribution of Component Maintenance Time

Expert-Based Mean	**Lognormal** Lower Bound	**Lognormal** Upper bound	**Data-Based** Mean
116.0	66.2	467.0	265
40.4	23.1	162.6	29
20.9	11.9	84.1	11
10.8	6.2	43.5	7
116.0	66.2	467.0	135
40.4	23.1	162.6	19
20.9	11.9	84.1	4
116.0	66.2	467.0	580
116.0	66.2	467.0	39
40.4	23.1	162.6	37
20.9	11.9	84.1	14
10.8	6.2	43.5	6

6.2.2 Adult Weight Management Case Study

A weight management survey instrument (see Appendix E) was administered to registered dieticians with varying degrees of expertise. Adult weight management is a specialized sub-discipline in the field of dietetics. Experts were given a clinical nutrition diagnostic problem regarding the ideal recommended "very low calorie diet" for an obese girl. Experts were asked to make a judgment about maximum recommended Kcal per day. Table 12 presents the estimates of experts and subsequent lognormal adjustment of their estimates.

Table 14. Expert Estimate and Lognormal Model Projections of the True Value of a Very Low Calorie Diet

Expert	Expert Estimate	Adjusted Expert Estimate: Lower Bound	Adjusted Expert Estimate: Upper Bound	Absolute Error of Lower bound
A	**600**	**342.5**	**2415.5**	0.57
B	**1200**	**685.1**	**4831.0**	0.14
C	**1200**	**685.1**	**4831.0**	0.14
D	**-------**	**-------**	**-------**	-------
E	**1500**	**856.4**	**6038.7**	0.07
F	**1000**	**570.9**	**4025.8**	0.29
Mean	**1125**	**628.0**	**4428.4**	**0.24**

True value of Kcal=800

6.3 VALIDATION OF EQUATIONS BY CASE STUDIES

In this section the regression equations developed in Chapter 5 to predict expert performance are validated. Given the attributes of an expert, these equations are designed to reasonably predict expert performance. Each equation for expert performance is a function of the following attributes: peer nominations, certification or specialized training in expertise, publications expertise or field, membership in professional organizations, and organization specialization in expertise, as well as institution type, average level of formal education, event frequency, and average years of experience. Data obtained from the Precipitation Forecasting and Human Nutrition case studies are used to evaluate the validity of Equations.

6.3.1 Precipitation Forecasting Case Study

A weather precipitation case study among expert meteorologists at the University of Maryland- College Park was performed. The objectives of the study were three-fold: (1) to predict the APE of experts given their estimates and (2) to determined the effect of expertise on expert performance. The third objective of the study will be discussed in the next chapter.

The case study involved four experts who were asked to make 48 hour precipitation forecasts projections. In the field of meteorology, a 48hour forecast of precipitation is considered moderately difficult, and requires specialized skills. The forecast were conducted on three different days for the following cities: Orlando, Seattle, San Francisco, New Orleans and Detroit. The complete record of the survey instrument is documented in Appendix E.

Findings from the case studies indicate that the regression equation predicted expert performance values comparable to the actual performance of the expert.

Table 15. Cumulative Attributes of Experts in the Precipitation Forecasting Case Study

EXPERTS	$\sum_{j=1}^{N} x_{ij} = X_i$	**Actual % Accuracy**	**Predicted % Accuracy** $p_i \propto \beta\, ln\lvert X_i\rvert$
1	5	60.0%	67.7%
2	6	66.7%	72.8%
3	7	66.7%	77.1%
4	9	66.7%	84.1%

Table 16. Weighted Attributes of Experts in the Precipitation Forecasting Case Study

EXPERTS	$\sum_{j=1}^{N} x_{ij} c_{ij} = X_i^C$	**Actual % Accuracy**	**Predicted % Accuracy**	
			$ln\lvert p_i\rvert \propto \beta\, ln\lvert X_i^C\rvert$	$p_i \propto \beta\, ln\lvert X_i^C\rvert$
1	2.24	60.0%	61.1%	64.9%
2	2.67	66.7%	70.3%	72.9%
3	2.34	66.7%	73.7%	75.5%
4	3.45	66.7%	80.9%	80.9%

6.3.1 Adult Weight Management Case Study

Adult overweight and obesity have become a public health concern as over 60% of the population has been reported to be overweight and 20% obese. Both conditions are predisposing factors for the development of chronic illnesses. Physicians, dieticians, nurses and exercise physiologists working in the area of adult weight management posses a common body of knowledge about the identification, assessment and treatment of obesity. Utilizing evidence-based information, adult weight management professionals sharpened their expertise by improving their skills and competence through specialized training in Adult weight management.

The case study entails, the elicitation of experts responses to 11 questions (see Appendix E-3) on Adult Weight Management, and the completion of a brief inquiry about your expertise. The identity of all experts remained anonymous. The cumulative attributes of experts in this case study was used to predict the performance of experts. The complete record of all expert performances in the study is documented in Appendix E-1 through E-3.

Table 17. Cumulative Attributes of Experts in the Adult Weight Management Case Study

EXPERTS	$\sum_{j=1}^{N} x_{ij} = X_i$	Actual % Accuracy	Predicted % Accuracy $p_i \propto \beta\, ln\lvert X_i \rvert$
A	6	91.0%	72.8%
B	7	72.7%	77.1%
C	6	63.6%	72.8%
D	7	63.6%	77.1%
E	7	72.7%	72.8%
F	5	72.7%	67.7%

Table 18. Weighted Attributes of Experts in the Adult Weight Management Case Study

EXPERTS	$\sum_{j=1}^{N} x_{ij} c_{ij} = X_i^C$	Actual % Accuracy	Predicted % Accuracy	
			$ln\lvert p_i \rvert \propto \beta\, ln\lvert X_i^C \rvert$	$p_i \propto \beta\, ln\lvert X_i^C \rvert$
A	1.01	91.0%	66.0%	69.3%
B	1.36	72.7%	74.3%	76.0%
C	1.12	63.6%	68.8%	71.6%
D	1.56	63.6%	78.5%	79.1%
E	1.12	72.7%	68.8%	71.6%
F	1.01	72.7%	66.0%	69.3%

Amongst the three regression relationships in Equations 5.4.3, 5.4.5 and 5.4.6, for predicting expert performance, the most effective was Equation 5.4.6:

$$ln\, p_i = \beta\, ln \left| \sum_{j=1}^{N} x_{ij} c_{ij} \right| + T$$

Among the participating experts in this component of the validation phase, the model above effectively forecasted the accuracy of half the experts within approximately five percent of their actual performance. In addition, the model also predicted the accuracy of three-fourth of the experts within 15 percent of their actual performance.

CHAPTER 7

CONCLUSIONS

The Quality of Expert Judgment is of value to decision makers for a variety of reasons. Decision makers need to be informed of the quality of their experts' judgments prior to implementation. The current work contains significant contributions to the expert judgment quality body of knowledge. This research presents a clear-cut approach to evaluate the performance of experts, and the true value of an unknown quantity.

Results suggest given an expert estimate about a given quantity, the lognormal model in Chapter 4 can reasonably predict/forecast the true value of that unknown quantity. In all four case studies, the lognormal model improved 60% to 75% of the experts' estimates. The estimates for the true values of the unknown quantities were calculated from the most likely value of *"u"* or the unknown quantity; use of the calibration posterior distribution is expected to improve estimates.

Another significant finding is the relationship of attributes to performance. The non-linear logarithmic regression equation in Chapter 5 appropriately illustrates this association. In this model, the performance of probability of experts correctly providing judgments is the dependent variable and the attributes are the explanatory variables. This implies, given the qualifications or attributes of an expert, performance can be predicted. The regression equation allows for the inclusion other attributes not evaluated in this study. Results from the validation case studies show that the empirical regression equation was effective in forecasting 50% of the elicited experts' ability to provide accurate responses within 5% of their actual performance. The model also predicted 75% if the experts' performance with 15% of actual scores.

<u>Recommendations:</u>

Continuing the fine-tuning of empirical relationships between *true values and expert estimates*, and *attributes and performance by:*

- Securing and applying equation of unknown quantity to additional expert judgment case studies
- Performing independent expert judgment case studies to adjust regression coefficients

<u>Implications:</u>

- Given an expert's estimate, the Bayesian equation for an unknown quantity can reasonably estimate the true value
- Given the attributes of an expert, his/her performance can be sufficiently predicted

REFERENCES

Afset, J. ; Lunde, P.; and Ramussen, K. (1996). Accuracy of routine echocardiographic measurements made by an inexperienced examiner through tele-instruction. *Journal of Telemedicine and Telecare*, 2, 148-154.

Armstrong, Scott J (1985). Long-Range Forecasting: From Crystal Ball to Computer. John Wiley & Sons, New York, pp73.

Ascher, W (1978). Forecasting: An Appraisal for Policymakers and Planners. Baltimore, Maryland, Johns Hopkins University Press.

Bariciak, E.; Plint, A.; Gaboury, I.; and Bennett, S. (2003). Dating of Bruises in Children: An Assessment of Physician Accuracy. *Pediatrics,* 112; 804-807.

Bentley, E., Jenkins., Campbell, F., and Warren, B. (2002). How could pathologists improve the initial diagnosis of colitis? Evidence from an international workshop. *Journal of Clinical pathology*, 55, 955-960.

Booker J.; Meyer M (1988). Sources and Effects of Inter-Expert Correlation: An Empirical Study. IEEE Transactions on Systems, Management, and Cybernetics, 8: 135-142.

Booker J.; Meyer M (1996). Elicitation and Analysis of Expert Judgment. Los Alamos National Laboratory, LA-UR-99-1659.

Brancato B.; Bricolo D.; Bonzanini M.; Ciatto S.; Mariotto R.; Miccinesi G.; Zappa M (2002). Pre-Operative Prediction of Invasive vs Intraductal Breast Cancer Type: Multivariate Analysis of the Accuracy of Clinical and Imaging Findings. The Breast, 11(2): 151-155.

Brown E.; Bruce M.; McAvay G.; Raue P.; Lachs M.; and Nassisi P. (Jun 2004). Recognition of Late-Life Depression in Home Care: Accuracy of the Outcome and Assessment Information Set. *Journal of the American Geriatrics Society,* 52(6); 995-999.

Bruynesteyn, K.; Van Der Heijde, D.; Boers, M.; Saudan, A.; Peloso, P.; Paulus, H.; Houben, H.; Griffiths, B.; Edmonds, J.; Bresnihan, B.; Boonen, A.; and Van Der Linden, S. (2002). Detecting Radiological changes in Rheumatoid Arthritis That Are Considered Important by Clinical Experts: Influence of Reading With or Without Known Sequence. *The Journal of Rheumatology*, 29(11); 2306-2312.

Budescu, D.V. and Rantilla, A.K. (2000). Confidence in aggregation of expert opinions. Acta Psychologica, 104, 371-398.

Camerer, C.; Johnson, E (1997). The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In W.M. Goldstein, and R.M. Hogarth (Eds.), Research on judgment and decision making, pp: 343-344, 347-348. Cambridge: Cambridge University Press.

Cook, R.M (1991). Experts in Uncertainty: Opinion and Subjective Probability in Science. Oxford Press, New York.

Cox, K (2002). Perceiving Clinical Evidence. Medical Education, 36: 1189-1195.

Dalkey, N.C (1969). The Delphi Method: An Experimental Study of Group Opinion, prepared for the United States Air Force Project RAND. Rand Corporation, Santa Monica, California.

Daneshkhah, A (2004). Uncertatinty in Probabilitstic Risk Assessment : A Review. http://www.shef.ac.uk/content/1/c6/03/09/33/risk.pdf.

De Finetti, B (1964). "La pervision: ses loislogique, ses source subjectives," Annales de Linstut Henri Poincare,7: 1-68 (1937). English translation in H. Kyburg and H Smokler (eds.), Studies in Subjective Probability, Wiley, New York.

Dreyfus, H.; Dreyfus, S (1986). Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer. New York: MacMillan.

Dreyfus, H.; Dreyfus, S (1996). The Relationship of Theory and Practice in the Acquisition of Skill. In Benner, P.; Tanner, C.; Chesla, C.; (eds.). Expertise in Nursing Practice (pp: 29-47). New York, Springer Publishing Company, Inc..

Dunphy, B., and Williamson, S. (2004). In Pursuit of Expertise: Toward an Educational Model for Expertise Development. *Advances in Health Sciences Education, 9;* 107-127.

French, S (1985). Group Consensus Probability Distribution: A Critical Survey. In Bayesian Statistics 2nd ed. JM Bernado et al. North Holland, Amsterdam, pp 183-201.

Genest, C.; Zidek, J (1986). Combining Probability Distribution: A Critique and an Annotated Biliography. Statistical Science, 1: 114-148.

Gobalet, K (2001). A Critique of Faunal Analysis; Inconsistency Among Experts in Blind Tests. Journal of Archaeological Science, 28(4): 377-386.

Haber, L.; Haber, R (2003). Error Rates for Human Latent Fingerprint Examiners. Human Factors Consultants and the University of California at Santa Cruz. Advances in Automatic Fingerprint. Springer-Verlag. New York.

Hughes, A.; Daniel, S.; Kilford, L.; and Lees, A. (1992). Accuracy of clinical diagnosis of idiolopa Parkinson's disease. A Clinico-pathological study of 700 cases. *J Neurol Neurosurg Psychiatry, 55;*181-184.

Jankovic, J.; Rajput, A.; McDermott, M.; and Perl D. (2000). The Evolution of Diagnosis in Early Parkinson Disease. *Arch Neurol, 57;* 371.

Jenkins, R (2003). Survey of Various Elicitation Methodologies for Expert Judgment in Risk Analysis. University of Maryland-College Park. Reliability Engineering Program (Special Project).

Kane, R. (1995). Creating Guidlelines: The Dangers of Over-Reliance on Expert Judgment. *Journal of Law, Medicine & Ethics,* 23; 62-64.

Lee, S (2002). Optimal Consensus of Fuzzy Opinions Under Group Decision Making Environment. Fuzzy Sets and Systems, 132(3): 303-315.

Lipinski, M.; Froelicher, V.; Atwood, E.; Tseitlin, A.; Franklin, B.; Osterberg, L.; and Myers, J. (2002). Comparison of treadmill scores with physician estimates of diagnosis and prognosis in patients with coronary artery disease. *American Heart Journal, 143(4);* 650-658.

Margo, C (2000). Peer and Expert Opinion and the Reliability of Implicit Case Review. Ophthalmology, 109(3): 614-618.

Meyer, M.; Booker, J (1991). Eliciting and Analyzing Expert Judgment: A Practical Guide, Academic Press, London, UK.

Moon, J.; Kang, C (1999). Use of Fuzzy Set Theory in the Aggregation of Expert Judgments. Annals of nuclear Energy, 26(6): 461-469.

Mosleh, A (1981). On the Use of Quantitative Judgment in Risk Assessment: A Bayesian Approach. Dissertation, University of California, Los Angeles.

Mosleh, A (1992). Bayesian Modeling of Expert-to-Expert Variability and Dependence in Estimating Rare Event Frequencies. Reliability Engineering and System Safety, 38.

Mrus, J (2004). Getting Beyond Accuacy: Moving Toward Approaches That Can be Used in Practice. Clinical Infectious Disease, 38: 1391-1393.

Mumpower, J, and Stewart, T. (1996). Expert Judgment and Expert Disagreement. *Thinking and Reasoning,* 2; 191-211.

Penrod, S., and Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy and Law,* 1; 817-845.

Penrod, S.D., Fulero, S., and Cutler, B. (1995). Expert psychological testimony on eyewitness reliability before and after Daubert. *Behavioral Sciences and the Law,* 13; 229-260.

Ramsey, F. "Truth and Probability" in Braithwaithe (Ed.). The Foundations of Mathematics, Kegan Paul. London (1931) pp156-198.

Reischman, R and Yarandi, H. (2002). Critical Care Cardiovascular Nurse Expert and Novice Diagnostic Cue Utilization. *Journal of Advanced Nursing,* 39(1); 24-34.

Roebber, P., and Bosart, L. (1996). The Complex Relationship between Forecast Skill and Forecast Balue: A Real-World Analysis. *Weather and Forecasting,* 11; 544-559.

Rowe, G., and Wright, G. (2001). Differences in Expert and Lay Judgments of Risk: Myth or Reality? *Risk Analysis,* 21(2); 341-356.

Santori (2004). Preliminary results of an expert-opinion elicitation process to prioritize an informative system funded by Italian Ministry of Health for cadaveric donor management, organ allocation, and transplantation activity. Transplantation proceedings [0041-1345] 36:433 -434.

Sboner, A.; Bauer, P.; Zumiani, G.; Eccher, C.; Blanzier, E.; Forti, S.; and Cristofolini, M. (2004). Clinical Validation of an Automated System for Supporting the Early Diagnosis of Melanoma. *Skin Research and Technology: Official Journal of International Society for Bioengineering and the Skin (ISBS) and International Society for Digital Imaging of Skin (ISDIS) and International Society for Skin Imaging (ISSI),* 10(3); 184-192.

Shanteau, J. (1995). Expert Judgment and Financial Decision Making. Prepared for: *Risky Business: Risk Behavior and Risk Management,* Bo Green (Ed.). Stockholm University.

Slovic, P. (1999). Trust, emotion, sex, politics and science: Surveying the risk-assessment battlefield. *Risk Analysis,* 19; 689-701.

Sohn, K.; Yang, J.; Kang, C (2001). Assimilation of Public Opinions in Nuclear Decision-Making Using Risk Perception. Annals of Nuclear Energy, 38: 553-563.

Sorrento, A; and Pichichero, M. (2001). Assessing Diagnostic Accuracy and Tympanocentesis Skills by Nurse Practitioners in Management of Otitis Media. *Journal of the American Academy of Nurse Practitioners,* 13(11), 524-529.

Stewart, T (2000). Uncertainty, Judgment, and Error in Prediction. In, Prediction: Science, Decision Making, and the Future of Nature, edited by Daniel Sarewitz, Roger A. Pielke, Jr., and Radford Byerly, Jr. Island Press, Washington DC.

Vegelin, A.; Brukx, L.; Waelkens, J.; and Van den Broeck, J. (2003). Influence of knowledge, training and experience of observers on the reliability of anthropometric measurements in children. *Annals of Human Biology*, 30(1); 65-79.

Weiss, D., and Shanteau, J. Do Judgments Alone Provide Sufficient Information to Determine the Expertise of the Judge Who Made Them? (unpublished)

Weiss, D., and Shanteau, J. (2003). Empirical Assessment of Expertise. *Human Factors*, 45(1); 104-114.

Wright, G.; Bolger, F (1992). Introduction. In G. Wright and F. Bolger (eds.), Expertise and Decision Support (pp1-8). New York: Plenum.

REFERENCES: CASE STUDIES

Armstrong, J (1983). Relative Accuracy of Judgmental and Extrapolative Methods in Forecasting Annual Earnings. Journal of Forecasting, 2: 437-447.

Barefield, R.; Comiskey, E.; McDonald, C (1979). Accuracy of Management and Security Analysts' Forecasts: Additional Evidence. Journal of Business Research, 7: 109-115.

Bariciak, E.; Plint, A.; Gaboury, I.; and Bennett, S (2003). Dating of Bruises in Children: An Assessment of Physician Accuracy. Pediatrics, 112: 804-807.

Basi, B.; Carey, K.; Twark, R (1976). A Comparison of the Accuracy of Corporate and Security Analysts' Forecasts of Earnings. The Accounting Review, 51(2): 244-254.

Bentley, E.; Jenkins, D.; Campbell, F.; Warren, B (2002). How Could Pathologists Improve the Initial Diagnosis of Colitis? Evidence from an International Workshop. Journal of Clinical Pathology, 55: 955-960.

Benke, G.; Sim, M.; Forbes, A.; Salzberg, M (1997). Retrospective Assessment of Occupational Exposure of Chemicals in Community-Based Studies. Validity and Repeatability of Industrial Hygiene Panel Ratings. International Journal of Epidemiology, 26: 635-642.

Bowers, W.; Jarrell, S (Summer 1992). Accuracy Analysis of Consensus Forecasts. The Journal of Business Forecasting Methods and Systems, 11.

Brandon, C.; Jarrett, J (1977). Accuracy of Externally Prepared Forecasts. Review of Business and Economic Research, 13: 35-47.

Brancato, B.; Bricolo, D.; Bonzanini, M.; Ciatto, S.; Mariotto, R.; Miccinesi, G.; Zappa, M (2002). Pre-Operative Prediction of Invasive vs Intraductal Breast Cancer Type: Multivariate Analysis of the Accuracy of Clinical and Imaging Findings. The Breast, 11(2): 151-155.

Brown, E.; Bruce, M.; McAvay, G.; Raue, P.; Lachs, M.; and Nassisi, P (Jun 2004). Recognition of Late-Life Depression in Home Care: Accuracy of the Outcome and Assessment Information Set. Journal of the American Geriatrics Society, 52(6): 995-999.

Bruynesteyn, K.; Van Der Heijde, D.; Boers, M.; Saudan, A.; Peloso, P.; Paulus, H.; Houben, H.; Griffiths, B.; Edmounds, J.; Bresnihan, B.; Boonen A.; Van Der Linden, S (2002). Detecting Radiological Changes in Rheumatoid Arthritis that are Considered Important by Clinical Experts: Influence of Reading With or Without Known Sequence. The Journal of Rheumatology, 29(11): 2306-2311.

Bustamante, M.; Devesa, F.; Borghol,A.; Ortuno, J.; Ferrando, M (2002). Accuracy of the Initial Endoscopic Diagnosis in the Discrimination of Gastric Ulcers. Journal of Clinical Gastroentrology, 35(1): 25-28.

Camerer, C.; Johnson, E. The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In W.M. Goldstein, and R.M. Hogarth (Eds.), Research on judgment and decision making. Cambridge: Cambridge University Press.

Catron, A (Summer 2000). Daily Demand Forecasting At Columbia Gas. The Journal of Business Forecasting, 10.

Collins, W.; Hopwood, W (1980). A Multivariate Analysis of Annual Earnings Forecasts Generated from quarterly Forecasts of Financial Analyst and Univariate Time-Series Models. Journal of Accounting Research, 18: 390-406.

Copeland, R.; Marioni, R (1972). Executives Forecasts of Earnings Per Share Versus Forecasts of Naïve Models. Journal of Business, 45: 497-512.

Cregan, P.; Parer, J.; Power, A (1988). Accuracy of Mammography in an Australian Community Setting. The Medical Journal of Australia, 149: 408-409.

Critchfield, T.; Dyckman, T.; and Lakonishok, J (1978). An Evaluation of Security Analysts' Forecasts. Accounting Review, 53: 651-668.

Culverhouse, P.; Williams, R.; Reguera, B.; Herry, V; Gonzalez-Gil, S (2003). Do Experts Make Mistakes? A Comparison of Human and Machine Identification of Dinoflagellates. Marine Ecology Progress Series, 247: 17-25.

Della Mea, V.; Forti, S.; Puglisi, F. ; Bellutta, P. ; Finato, N. ; Dalla Palma,D. ; Mauri, F. ; Beltrami, C (1996). Telepathology using Interanet Multimedia Electronic Mail : Remote Consultation on Gastrointestinal Pathology. Journal of Telemedicine and Telecare, 2: 28-34.

Donatelli, J.; Macone, A.; Goldman, D.; Poon, R.; Hinberg, I.; Nanji, A.; and Thorne, G (Jan 1992). Rapid Detection of Group A Stretococci: Comparative Performance by Nurses and Laboratory Technologists in Pediatric Satellite Laboratories Using Three Test Kits. Journal of Clinical Microbiology, 30(1): 138-142.

Dougall, N.; Nobili, F.; Ebmeier, K (2004). Predicting the Accuracy of Diagnosis of Alzheimer's Disease with ^{99m}Tc HMPAO Single Photon Emission Computed Tomography. Psychiatry Research: Neuroimaging.

Dowse, G. The Accuracy of Qualitative Residential Market Forecasts: the Auckland Experience. Department of Finance, Banking and Property, Massey University Palmerston North, New Zealand, unpublished.

Ekstrom, G.; Johnsson, T.; Borrman, H (1993). Accuracy Among Dentists Experienced in Forensic Odontology in Establishing Identity. The Journal of Forensic Odonto-Stomatology, 11(2): 45-53.

Elton, E.; Gruber, M (1972). Earnings Estimates and the Accuracy of Expectational Data. Management Science, 18, B409-B423.

Engstrom, J.; McFarlin, B.; and Sampson, M (Dec 1993). Fundal Height Measurement: Part4—Accuracy of Clinicians' Identification of the Uterine Fundus During Pregnancy. Journal of Nurse-Midwifery, 38(6): 318-323.

Fried, D.; Givoly, D (1982). Fiancial Analysts Forecasts of Earnings: A Better Surrogate for Market Expectations. Journal of Accounting and Economics, 4: 85-107.

Frijling, B.; Spies, T.; Lobo, C.; Hulscher, M.; Van Drenth, B.; Braspenning, J.; Prins, A.; Van der Wouden, J.; and Grol, R (2001). Blood pressure control in treated hypertensive patients: clinical performance of general practitioners. British Journal of General Practive, 51: 9-14.

Fritsch, L.; Nadon, L.; Geza Benke.; Lakhani, R.; Latreille, B.; Parent, M. ; Siemiatycki, J (2003). Validation of Expert Assessment of Occupational Eposures. American Journal of Industrial Medicine, 43: 519-522.

Galant, S.; Crawford, L.; Morphew, T.; Jones, C.; and Bassin S (2004). Predictive Value of a Cross-Cultural Asthma Case-Detection Tool in an Elementary School Population. Pediatrics, 114(3): 307-316.

Hagerman, R.; Ruland, W (1979). The Accuracy of Management Forecasts and Forecasts of Simple Alternative Models. Journal of Economics and Business, 31: 172-179.

Hatala, R.; Norman, G.; Brooks, L (1999). Impact of a Clinical Scenario on Accuracy of Electrocardiogram Interpretation. Journal of General Internal Medicine, 14: 126-139.

Hughes, A.; Daniel, S.; Ben-Shlomo, Y.; Lees, A (2002). The Accuracy of Diagnosis of Parkinsonian Syndromes in a Specialist Movement Disorder Service. Brain, 125: 861-870.

Hughes, A.; Daniel, S.; Kilford, L.; Lees, A (1992). Accuracy of Clinical Diagnosis of Idiopathic Parkinson's Disease: a clinio-pathological study of 100 cases. Journal of Neurology Neurosurgery Psychiatry, 55: 181-184.

Imhoff, E (1978). The Representativeness of Management Earnings Forecasts. The Accounting Review, 53(4): 836-850.

Jaggi, B (1978). Comparative Accuracy of Management's Annual Earnings Forecasts. Financial Management, 7: 24-32.

Jankovic, J.; Rajput, A.; McDermott, M.; Perl, D (2000). The Evolution of Diagnosis in Early Parkinson Disease. Arch Neurol, 57: 369-372.

Jayes, R.; Larsen, G.; Beshansky, J.; D'Agostino, R.; and Selker, H (July/August 1992). Physician Electrocardiogram Reading in the Emergency Department—Accuracy and Effect on Triage Decisions. Journal of General Internal Medicine, 7: 387-392.

Jenkins, M.; Clarke, J.; Carlin, J.; Robertson, C.; Hopper, J.; Dalton, M.; Holst, D.; Choi, K.; and Giles, G (1996). Validation of Questionnaire and Bronchial Hyperresponsiveness against Respiratory Physician Assessment in the Diagnosis of Asthma. International Journal of Epidemiology, 25(3): 609-616.

Kaleida, P.; and Stool S (April 1992). Assessment of Otoscopists' Accuracy Regarding Middle-Ear Effusion. AJDC, 146: 433-435.

Kaplan, A.; Skogstad, A.; and Girshick, M (Spring 1950). The Prediction of Social and Technological Events. Public Opinion Quarterly, 93-110.

Keilman, N (2001). Data Quality and Accuracy of United Nations Population Projections, 1950-95. Population Studies, 55: 149-164.

Keilman, N (1997). Ex-post Errors in Official Population Forecasts in Industrialized Countries. Journal of Official Statistics, 13:245-277.

Keyfitz, N (1982). Can knowledge Improve Forecasts? Population and Development Review, 8: 729-751.

Lai, K.; Pauly, P. Random Walk or Bandwagon: Some evidence from foreign exchanges in the 1980s. Applied Economics, 24, 693-700.

Ledesert, B.; and Ritchie, K (1994). The Diagnosis and Management of Senile Dementia in General Practice: A Study of 301 General Practitioners in the Monpellier Region. International Journal of Geriatric Psychiatry, 9: 43-46.

Leggett, P.; Graham, L.; Steele, K.; Gilliland, A.; Stevenson, M.; O'Reilly, D.; Wootton, R.; and Taggart, A (Sep 2001). Telerheumatology—Diagnostic Accuracy and Acceptability to Patient, Specialist, and General Practitioner. The British Journal of General Practice: the Journal of the Royal College of General Practitioners, 51(470): 746-748.

Levinthal, G.; Burke, C.; and Santisi, J (Dec 2003). The Accuracy of an Endoscopy Nurse in Interpreting Capsule Endoscopy. The American Journal of Gastroenterology, 98(12): 2669-2671.

Litvan, I.; Goetz, C.; Jankovic, J.; Wenning, G. ;Booth, V.; Bartko, J (1997). What is the Accuracy of the Clinical Diagnosis of Multiple System Atrophy? Arch Neurol, 54: 937-944.

Lipinski, M.; Froelicher, V.; Atwood, E.; Tseitlin, A.; Franklin, B.; Osterberg, L.; Do, D.; Myers, J (2002). Comparison of Treadmill Scores with Physician Estimates of Diagnosis and Prognosis in Patients with Coronary Artery Disease. American Heart Journal, 143(4): 650-658.

Ljung, B.; Drejet, A.; Chiampi, N.; Jeffrey, J.; Goodson, W.; Chew, K.; Moore, D.; and Miller, T (Aug 2001). Diagnostic Accuracy of Fine-Needle Aspiration Biopsy is Determined by Physician Training in Sampling Technique. Cancer, 93(4): 263-268.

Llanos, O.; Guzman, S.; Duarte, I (1982). Accuracy of the First Endoscopic Procedure in the Differential Diagnosis of Gastric Lesions. Ann Surg, 195(2): 224-226.

Lorek, K.; McDonald, C.; Patz, D (1976). A Comparative Examination of Management Forecasts and Box-JenkinsForecasts of Earnings. Accounting Review, 51: 321-330.

Luck, J.; Peabody (2002). Using Standardised Patients to Measure Physicians' Practice: Validation Study Using Audio Recordings. BMJ, 325.

Makridakas, S.; Winkler, R (September 1983). Averages of Forecasts: Some Empirical Results. Management Science, 29: 987-996.

MacKenzie-Wood, A.; Milton, G.; Launey, J (1988). Melanoma: Accuracy of Clinical Diagnosis. Australasian Journal of Dermatology, 39: 31-33.

Maniatis, A.; Eisen, G.; Braxer, S (1997). Endoscopic Discrimination of Gastric Ulcers. Journal of Clinical Gastroenterology, 24(4): 203-206.

McClish, D.; Powell, S (1989). How Well Can Physicians Estimate Mortality in a Medical Intensive Care Unit? Medical Decision Making, 9: 125-132.

McNees, S (July/August 1988). How Accurate Are Macroeconomic Forecasts? New England Economic Review. pp15-36.

Messite, J.; and Stellman, S (March 1996). Accuracy of Death Certificate Completion. JAMA, 275(10): 794-796.

Mills, J.; Webster, A.; Wofsy, C.; Harding, P.; and D'Acuti, D (Nov 1976). Effectiveness of Nurse Triage in the Emergency Department of an Urban County Hospital. JACEP, 5: 877-882.

Morgan, T.; Civil, I.; and Schwab C (May 1988). Injury Severity Scoring: Influence of Timing and Nurse Raters on Accuracy. Heart Lung, 17(3): 256-261.

Motherby, H.; Nadjari, B.; Friegel, P.; Kohaus, J.; Ramp, U.; and Bocking A (1999). Diagnostic Accuracy of Effusion Cytology, 20(6): 350-357.

Muram, D.; Arheart, K.; Jennings, S (1999). Diagnostic Accuracy of Colposcopic Photographs in Child Sexual Abuse Evaluations. J Pediatr Adolesc Gynecol, 12: 58-61.

Neale, A; Demers, R; Budev, H.; and Scott, R (April 1987). Physician Accuracy in Diagnosing Colorectal Polyps. Dis Colon Rectum, 30(4): 247-250.

Nordrum, I.; Johansen, M.; Amin, A.; Isaksen, V. ; Ludvigsen, J (2004). Diagnostic Accuracy of Second-Opinion Diagnoses Based on Still Images. Human Pathology, 35(1): 129-135.

Onkal, D.; Yates, J.; Simga-Mugan, C.; and Oztin S (2003). Professional vs. amateur judgment accuracy: The case of foreign exchange rates. Organizational Behavior and Human Decision Processes, 91: 169-185.

Peabody, J.; Luck, J.; Glassman, P.; Jain, S.; Hansen, J.; Spell, M.; and Lee, M (Nov 2004). Measuring the Quality of Physician Practice by Using Clinical Vignettes: A Prospective Validation Study. Annals of Internal Medicine, 141(10): 771-780.

Pederson, L; and Baskerville, J (1983). Multivariate Prediction of Smoking Cessation following Physician Advice to Quit Smoking: A Validation Study. Preventive Medicine 12: 430-436.

Picardi, M.; Gennarelli, N.; Ciancia, R.; et al. (2004). Randomized Comparison of Power Doppler Ultrasound-Directed Excisional Biopsy with Standard Exciional Biopsy for the Characterization of Lymphadenopathies in Patients with Suspected Lymphoma. J Clin Oncol, 22: 3733-3740.

Piccolo, D.; Soyer, P.; Burgdorf, W.; Talamini, R.; Peris, K.; et al (2002). Concordance Between Telepathologic Diagnosis and Conventional Histopathologic Diagnosis : A Multiobserver Store-and-Forward Study on 20 Skin Specimens. Arch Dermatol, 138: 53-58.

Pichichero, M (2002). Diagnostic Accuracy, Tympanocentesis Training Performance, and Antibiotic Selection by Pediatric Residents in Management of Ottis Media. Pediatrics, 110: 1064-1070.

Quigley, M.; Chandramohan D.; and Rodrigues, L (1999). Diagnostic Accuracy of Physician Review, Expert Algorithms and Data-derived Algorithms in Adult Verbal Autopsies. International Journal of Epidemiology, 28: 1081-1087.

Raab, S.; Oweity, T.; Hughes, J.; Salomao, D.; Kelly, C.; Flynn, C.; D'Antonio, J.; Cohen, M (2000). Effect of Clinical History on Diagnostic Accuracy in the Cytologic Interpretation of Bronchial Brush Specimens. Am J Clin Path, 114: 78-83.

Ranazzo, M.; Snoey, E.; Levitt, A.; and Binder, K (Sept 2003). Accuracy of Emergency Physician Assessment of Left Ventricular Ejection Fraction and Central Venous Pressure Using Echocardiography. Academy of Emergency Medicine, 10(9): 973-977,.

Rajput, A.; Rozdilsky, B. ; Rajput, A (1991). Accuracy of Clinical Diagnosis in Parkinsonism: A Prospective Study. Can J Neurol Sci, 18: 275-8.

Reischman, R (Jul 2002). Critical Care Cardiovascular Nurse Expert and Novice Diagnostic Cue Utilization. Journal of Advanced Nursing, 39(1), 24-34.

Remy, M.; Rikers, P.; Schmidt, H (2002). On the Constraints of Encapsulated Knowledge: Clinical Case Representations by Medical Experts and Subexperts. Cognition and Instruction, 20(1): 27-45.

Richards, R.; Benjamin, J.; Strawser, R (1977). An Examination of the Accuracy of Earnings Forecasts. Financial Management, 6: 78-86.

Ruland, W (1978). The Accuracy of Forecasts by Management and by Financial Analysts. Accounting Review, 53: 439-447.

Ryan, D (1994). Misdiagnosis in Dementia: Comparisons of Diagnostic Error Rate and Range of Hospital Investigation According to Medical Specialty. International Journal of Geriatric Psychiatry, 9: 141-147.

Santana, J.; and Martinez, R (April 1995). Accuracy of Emergency Physician Data Collection in Automobile Collisions. The Journal of Trauma, 38(4): 583-586.

Santana-Boado, C.; Candell-Riera,J.; Castell-Conesa, J.; Aguade-Bruix, S.; Garcia-Burillo, A.; Canela,T.; Gonzalez, J.; et al (1998). Diagnostic Accuracy of Technetium-99m-MIBI Myocardial SPECT in Women and Men. J Nucl Med, 39: 751-755.

Sboner, A.; Bauer, P.; Zumiani, G.; Eccher, C.; Blanzier, E.; Forti, S.; and Cristofolini, M (Aug 2004). Clinical Validation of an Automated System for Supporting the Early Diagnosis of Melanoma. Skin Research and Technology: Official Journal of International Society for Bioengineering and the Skin (ISBS) and International Society for Digital Imaging of Skin (ISDIS) and International Society for Skin Imaging (ISSI), 10(3): 184-192.

Schoenfeld, P.; Lipscomb, S.; Crook, J.; Dominguez, J.; Butler, J.; Holmes, L.; and Cruess, D.; and Rex, D (Aug 1999). Accuracy of Polyp Detection by Gastroenterologists and Nurse Endoscopists During Flexible Sigmoidoscopy: a Randomized Trial. Gastroenterology, 117(2): 312-318.

Seymour, J.; Saunders, P.; Wattis, J.; and Daly, L (1994). Evaluation of Early Dementia by a Trained Nurse. International Journal of Geriatric Psychiatry, 9: 37-42.

Sorrento, A; and Pichichero, M (Nov 2001). Assessing Diagnostic Accuracy and Tympanocentesis Skills by Nurse Practitioners in Management of Otitis Media. Journal of the American Academy of Nurse Practitioners, 13(11): 524-529.

Spigel, D. Improving Diagnositc Accuracy in Patients with Suspected Lymphoma. JCOM, 11(11): 691-692.

Steiner, R.; Vote, D.; Belk, K.; Scanga.; Wise, J.; Tatum, J.; and Smith, G (2003). Accuracy and Repeatability of Beef Carcass Longissimus Muscle Area Measurements. Journal of Animal Science, 81: 1980-1988.

Stewart, T.; Moninger, W.; Grassia, J.; Brady, R.; Merrem, F (1989). Analysis of Expert Judgement in a Hail Forecasting Experiment. Weather and Forecasting, 4: 24-34.

Stoto, M (1983). The Accuracy of Population Projections. Journal of the American Statistical Association, 78: 13-20.

Stritzel, F (2003). Patient's Informed Consent Prior to Implant-Prosthetic Treatment: A Retrospective Analysis of Expert Opinions. The International Journal of Oral and Maxillofacial Implants, 18: 433-439.

Stults, K.; Brown, D.; and Kerber, R (April 1986). Efficacy of an Automated External Defibrillator in the Management of Out-of-Hospital Cardiac Arrest: Validation of the Diagnostic Algorithm and Initial Clinical Experience in a Rural Environment. Circulation, 73(4): 701-709.

Tragardh, B.; Haglund, U (1985). Endoscopic Diagnosis of Gastric Ulcer. Acta Chir Scand,151: 37-41.

Van Klei, W.; Hennis, P.; Moen, J.; Kalkman, C.; and Moons, K (Oct 2004). The Accuracy of Trained Nurses in Pre-operatie Health Assessment: Results of the OPEN Study. Anaesthesia, 59(10): 971-978.

Vegelin, A.; Brukx, L.; Waelkens, J.; and Van den Broeck, J (2003). Influence of knowledge, training and experience of observers on the reliability of anthropometric measurements in children. Annals of Human Biology, 30(1): 65-79.

Verney, M.; Jelinger, K. Hauw, J.; Bancher, C.; Litvan, I.; Agid, Y (1996). Progressive Supranuclear Palsy : A Clinicopathological Study of 21 Cases. Acta Neuropathol, 91: 427-431.

Wenning, G.; Ben-Shlomo, Y.; Magalhaes, M.; Daniel, S.; Quinn, N (1995). Clinicopathological Study of 35 Cases of Multiple Sytem Atrophy. J Neurol Neurosurg Psychiatry, 58: 160-166.

Wildi, S.; Wallace, M.; Glenn, T.; Mokhashi, M.; Kim, C.; and Hawes, R (2003). Accuracy of esophagoscopy performed by a non-physician endoscopist with a 4-mm diameter battery-powered endoscope. Gastrointestinal Endoscopy, 57(3): 305-310.

Yaniv, I; Foster, D (1997). Precision and Accuracy of Judgmental Estimation. Journal of Behavioral Decision Making, 10: 21-32.

APPENDIX-A

SURVEY INSTRUMENT:
META-ANALYSIS ARTICLES DATA SHEET STUDY

SOURCE: ____

EXPERT ATTRIBUTES

Publication(s) on specific topic(s)	
Publication(s) in field	
Member of professional organization devoted to specific topic(s)	
Member of professional organization in field	
Average years of academic experience on specific topic(s):	
● 0 to <5 years	
● 5 to <10 years	
● >10 years	
Average years of academic experience in field:	
● 0 to <5 years	
● 5 to <10 years	
● >10 years	
Average years of practical experience on specific topic(s):	
● 0 to <5 years	
● 5 to <10 years	
● >10 years	
Average years of practical experience in field:	
● 0 to <5 years	
● 5 to <10 years	
● >10 years	
Nominated by peers as expert in field(s) or on specific topic(s)	
Number of experts in study	
Employee of a private company/ organization/ institution	
Employee of a public company/ organization/ institution	
Company/organization/institution specializes in specific/similar topic	
Other:	
Percent Accuracy:	
Percent Error:	
Specificity:	
Sensitivity:	

APPENDIX-B

LITERATURE-BASED EXPERT JUDGMENT CASE STUDIES ATTRIBUTES DATA SET

KEY

EXP	Number of Experts in the Study
NOM	Nominated by peers as expert in field(s) or expertise
INST	Type of Institution
EDU	Average level of formal education
CERT	Certified or received specialized training in expertise
PRCT	Average years of practical experience in expertise
PUB	Publication(s) in expertise or general field
MEM	Member of professional organization in expertise/field
COM	Company/organization specializes in specific/similar topic
EVNT	Frequency of event
ACC	Accuracy of Expert(s)

Data #	# EXP	NOM	INST	EDU	CERT	PRCT	PUB	MEM	COM	EVNT	ACC
1	23	0		1	0	0	0	1	0	3	23
2	3	1	2	0	1	3	1	1	0	2	40
3	30	0	2	2	0	2	0	1	0	3	48

Data #	# EXP	NOM	INST	EDU	CERT	PRCT	PUB	MEM	COM	EVNT	ACC
4	26	1	1	1	1	2	1	1	0	2	49.5
5	.	0	1	4	1	2	1	1	1	1	50
6	383	1	2	2	0	1	0	1	0	3	41
7		0		3	0	3	1	1	0	1	50
8	206	0	1	1	0	1	1	1	0	1	42
9	8	0	2	3	0	2	0	0	0	2	52.9
10	39	0	2	2	0	1	0	0	0	2	52.6
11	11	1	1	4	1	2	1	1	1	1	52.3
12	236	0	2	3	0	1	0	1	0	3	57
13	12	0	2	3	0	2	0	1	0	3	56.9
14	21	0	2	2	0	1	0	1	0	3	56
15	35	0	2	2	0	0	0	1	0	3	55.7
16		1	2	2	0	1	1	1	0	3	51
17	28	1		2	1	2	1	1	0	2	57
18	64	0	1	1	1	3	0	0	0	3	69.5
19	5	1	2	4	1	2	1	1	0	3	70.3
20	16	1	2	4	1	3	1	1	1	3	67
21	15	0	2	2	0	2	0	1	0	3	68
22	2	1	1	4	1	1	1	1	1	2	68.4
23	8	0	2	2	1	2	0	1	0	3	70.2
24	16	0	2	4	0	2	0	0	0	2	76.2
25	13	1	1	4	1	2	1	1	0	2	73.5
26	12	0	1	4	1	1	0	0	0	2	74
27	4	0	2	1	1	2	0	1	0	3	77.4
28		1	1	0	1	2	1	1	1	3	75
29	1	1	1	0	1	2	1	1	0	3	79.5
30	183	0	2	2	0	1	1	0	0	3	72.8

Data #	# EXP	NOM	INST	EDU	CERT	PRCT	PUB	MEM	COM	EVNT	ACC
31	12	0	2	3	0	2	1	1	0	3	80
32	2	1	1	4	1	2	1	1	0	1	71
33	45	1	2	4	1	2	1	1	0	3	76
34	37	0	1	4	0	2	1	1	0	3	73
35	29	0	1	2	0	2	1	0	0	3	76
36	13	0	1	4	0	2	0	1	0	3	71
37	27	0	1	4	0	2	0	1	0	3	80
38		0	1	4	1	3	1	1	1	1	72.1
39	3	1	1	4	1	2	1	1	0	3	77
40	116	0	1	3	0	2	1	1	0	3	73
41	16	1	1	4	1	2	1	1	1	1	78
42	5	0	2	4	1	2	1	1	0	3	80
43	23	1		1	1	2	1	1	0	3	72
44		1	1	4	1	2	1	1	0	3	80
45		1	1	3	0	2	1	1	0	3	76
46		0	2	3	0	2	0	1	0	3	74
47	8	1	1	4	1	2	1	1	0	1	74.5
48	4	0	1	4	1	1	1	1	0	2	71
49	151	0		4	1	2	1	1	0	1	73
50		1		4	1	2	1	1	0	3	85.8
51	6	0	1	4	1	2	1	1	1	2	83.7
52	4	0	2	4	1	2	1	1	1	2	89
53	4	0	1	1	1	2	0	1	0	3	80.6
54	11	1	1	4	1	2	1	1	1	1	89.8
55	11	1	1	4	1	2	1	1	1	1	87
56	11	1	1	4	1	2	1	1	1	1	88.7
57	15	0	2	2	1	2	0	1	0	3	81.5

Data #	# EXP	NOM	INST	EDU	CERT	PRCT	PUB	MEM	COM	EVNT	ACC
58	3	0	2	1	0	2	0	1	0	3	80.5
59	6	0	2	4	1	2	1	1	0	3	89
60	144	0	1	2	0	2	1	1	0	3	90
61	3	1	2	4	1	2	1	1	1	1	84.6
62	6	1		4	1	2	1	1	0	1	82
63	16	1	1	4	1	2	1	1	1	2	85
64		1	1	4	1	2	1	1	0	3	86
65		0	2	4	1	2	1	1	0	3	89.5
66	3	0	1	1	1	1	1	1	0	2	83
67	3	0	2	1	1	2	1	1	0	3	87
68	1	0	2	0	1	2	0	1	1	2	84.3
69	17	1	1	4	1	2	1	1	0	3	96.8
70	15	1	2	4	1	2	1	1	0	3	96
71	11	1	1	4	1	2	1	1	1	2	96
72	34	1	1	4	1	2	1	1	1	2	91.9
73	2	1	1	4	1	2	1	1	0	1	97
74	1	0	1	1	1	3	1	1	0	2	93
75	2	0		4	1	2	1	1	1	3	98.8
76		1	1	4	1	2	1	1	0	3	99
77	21	0	2	2	1	2	0	1	0	2	98
78	4	1		0	1	2	1	1	1	3	94
79		0	1	0	1	2	0	0	0	3	90.5
80		0	2	4	1	2	1	1	0	3	96

APPENDIX –C

SAMPLES OF ATTRIBUTES DESCRIPTIONS FROM CASE STUDIES

Author(s)	Expert Attributes	# Experts
Mitchell et. al., 1993	"Panel of psychiatric experts was constituted on the basis of peer nominators in 44 countries. It contained the names of potential respondents with extensive experience and knowledge of the pharmacotherapy of anxiety and depressive disorders. The final membership was selected on the basis of frequency of nomination by scientific peers, and encompassed 25 countries. Experts were finally selected based on the most nominations received"	66
Bentley et al, 2002	"An expert gastrointestinal pathologist was defined as someone having: membership of a professional organization devoted to gastrointestinal	13

Author(s)	Expert Attributes	# Experts
	pathology of inflammatory bowel disease; a diagnostic practice of at least 1000 gastrointestinal specimens each year; and at least five years of specialist gastrointestinal pathology."	
Bentley et al, 2002	"A general pathologist was defined as one: practicing in a community/ district general hospital practice; who had no nationally declared interest/ involvement in gastrointestinal pathology or publications, and whose involvement with gastrointestinal pathology was less than 40% of the workload"	12
Dougall et al, 2004	"Sixteen experts participated in the study. All had been responsible for clinical reporting of SPECT brain images within the last 10 years, with 12 describing themselves as Nuclear Medicine Specialists. Six experts used SPM in their current clinical practice. The 16 experts were recruited from several European sites."	16
Nordrum et al, 2004	"The referring pathologist (AA) from Department of Pathology, County	

Author(s)	Expert Attributes	# Experts
	Hospital of Nordland, has 25 years of pathology experience. This department has 5 consultants and 1 resident and handles 10,000 histological specimens, 21,000 cytological specimens, and 110 autopsies annually. The department has 10 consultants and 7 residents, and handles about 22,500 histological specimens, 35,000 cytological specimens, and 375 autopsies annually. Two general surgical pathologists (V.I. and I.N.) with 8 and 14 years of experience, respectively, diagnosed the cases in the study."	
Jankovic et al, 2000	"The 34 investigators at 28 centers were selected to participate in the study because they had a major interest in movement disorders and considerable experience in treating patient with Parkinson's Disease."	38
Hughes et al, 2002	Neurologists belonged to the National Hospital for Neurology and Neurosurgery in London	
Strietzel, 2003	"Expert belonged to the Dental Board of Berlin, Germany"	1
Litvan et al, 1998	"Experts in movement disorders"	

Author(s)	Expert Attributes	# Experts
Fritschi et al, 2003	"The raters undertaking this study were part of a team that has been involved in retrospective occupational exposure assessment in case-control studies of cancer in Montreal for nearly 20 years. The three chemists and industrial hygienists who carried out the rating for the present study had had an average of 10 years of experience between them. They also had access to a comprehensive library of information about exposures in workplaces."	4
Santori et al, 2004	"The panel was composed of a panel of nine independent experts, namely physicians and surgeons who had widely recognized competence at the international leveling the field of organ transplantation and related specialties, based on relevant experience, scientific production, and academic training."	9
Bruynesteyn et al, 2002	"Expert panel consisted of 5 rheumatologists (BB, BG, HH, HP, and PP) who independently evaluated 46 pairs of hand and foot films, taken at 1 year intervals, of patients with early RA with varying follow-up duration. The	5

Author(s)	Expert Attributes	# Experts
	experts were chosen from several different countries based on their expertise in the treatment of RA. None had been trained in either of the scoring method, but each was experienced in reading films daily practice."	

APPENDIX D-1

SURVEY INSTRUMENT:
EXPERT JUDGMENT IN WEATHER FORECASTING CASE STUDY

EXPERT ATTRIBUTES DATA SHEET

Please place an “X” or circle the appropriate response

1. Are you an employee of one of the following types of companies/ organizations/ institutions?

a. non-profit	b. private
b. academia	c. government
d. research/clinical	e. other

2. Do you have any peer-reviewed publication on forecasting precipitation?

c. Yes	b. No

3. Do you have any peer-reviewed publication on weather forecasting?

d. Yes	b. No

4. Have you been nominated by your peers as an expert in forecasting precipitation?

e. Yes	b. No

5. Have you been nominated by your peers as an expert in weather forecasting?

f. Yes	b. No

6. Are you a member of professional organization *specifically* devoted to forecasting precipitation?

g. Yes	b. No

7. Are you a member of professional organization devoted to forecasting weather?

h. Yes	b. No

8. Does your company/organization/institution specialize in forecasting precipitation?

i. Yes	b. No

9. What is your Highest Level of Formal Education?

j. Bachelor	b. Master
a. PhD	

10. How many years of experience (beyond formal education), do you have in the field of weather forecasting?

a. 0 to <5 years b. 5 to <10 years
c. >10 years

11. How many years of practical/field experience do you have in forecasting precipitation?

a. <1 year b. 1 to <5 years
c. 5 to <10 years d. >10 years
e. none

12. Do you have any specialized certification in forecasting precipitation?

k. Yes b. No

13. Do you have any specialized certification in forecasting weather?

l. Yes b. No

14. List/Describe any other expertise qualifications:

APPENDIX D-2

SURVEY INSTRUMENT:
EXPERT JUDGMENT IN WEATHER FORECASTING CASE STUDY

SAMPLE QUESTION SHEET

DAY 1

1. Will precipitation occur at the Orlando International Airport (MCO)?
   a. YES b. NO
   1b. If yes, how much? ___________________________

2. Will precipitation occur at the Seattle-Tacoma International Airport (SEA)?
   a. YES b. NO
   2b. If yes, how much? ___________________________

3. Will precipitation occur at the San Francisco International Airport (SFO)?
   a. YES b. NO
   3b. If yes, how much? ___________________________

4. Will precipitation occur at the Louis Armstrong New Orleans International Airport (MSY)?
   a. YES b. NO
   4b. If yes, how much? ___________________________

5. Will precipitation occur at the Detroit Metropolitan Wayne County Airport (DTW)?
   a. YES b. NO
   5b. If yes, how much? ___________________________

DAY 2

6. Will precipitation occur at the Orlando International Airport (MCO)?
   a. YES b. NO
   6b. If yes, how much? ___________________________

7. Will precipitation occur at the Seattle-Tacoma International Airport (SEA)?
   a. YES b. NO
   7b. If yes, how much? ___________________________

8. Will precipitation occur at the San Francisco International Airport (SFO)?
   a. YES b. NO

8b. If yes, how much? ____________________________

9. Will precipitation occur at the Louis Armstrong New Orleans International Airport (MSY)?
a. YES b. NO
9b. If yes, how much? ____________________________

10. Will precipitation occur at the Detroit Metropolitan Wayne County Airport (DTW)?
a. YES b. NO
10b. If yes, how much? ____________________________

<u>DAY 3</u>

11. Will precipitation occur at the Orlando International Airport (MCO)?
a. YES b. NO
1b. If yes, how much? ____________________________

12. Will precipitation occur at the Seattle-Tacoma International Airport (SEA)?
a. YES b. NO
2b. If yes, how much? ____________________________

13. Will precipitation occur at the San Francisco International Airport (SFO)?
a. YES b. NO
3b. If yes, how much? ____________________________

14. Will precipitation occur at the Louis Armstrong New Orleans International Airport (MSY)?
a. YES b. NO
4b. If yes, how much? ____________________________

15. Will precipitation occur at the Detroit Metropolitan Wayne County Airport (DTW)?
a. YES b. NO
5b. If yes, how much? ____________________________

APPENDIX D-3

SURVEY INSTRUMENT:
EXPERT JUDGMENT IN WEATHER FORECASTING CASE STUDY

PARTICIPANTS ATTRIBUTES DATA SHEET: FORECASTING CASE STUDY

Participant #1

Please place an "X" or **bold** the appropriate response

1. Are you an employee of one of the following types of companies/ organizations/ institutions?

m. non-profit	d. private
n. ***academia***	e. government
c. research/clinical	f. other

2. Do you have any peer-reviewed publication on forecasting precipitation?

a. Yes	**b.** ***No***

3. Do you have any peer-reviewed publication on weather forecasting?

a. Yes	**b.** ***No***

4. Have you been nominated by your peers as an expert in forecasting precipitation?

a. Yes	**b.** ***No***

5. Have you been nominated by your peers as an expert in weather forecasting?

a. Yes	**b.** ***No***

6. Are you a member of professional organization *specifically* devoted to forecasting precipitation?

a. ***Yes***	b. No

7. Are you a member of professional organization devoted to forecasting weather?

a. ***Yes***	b. No

8. Does your company/organization/institution specialize in forecasting precipitation?

a. ***Yes***	b. No

9. What is your Highest Level of Formal Education?

a. Bachelor	b. ***Master***
c. PhD	

10. How many years of experience (beyond formal education), do you have in the field of weather forecasting?

a. ***0 to <5 years*** b. 5 to <10 years
c. >10 years

11. How many years of practical/field experience do you have in forecasting precipitation?

a. ***<1 year*** b. 1 to <5 years
c. 5 to <10 years d. >10 years
e. none

12. Do you have any specialized certification in forecasting precipitation?

b. Yes b. ***No***

13. Do you have any specialized certification in forecasting weather?

c. Yes **b.** ***No***

14. List/Describe any other expertise qualifications:

PHD graduate student in Atmospheric and Ocean Sciences specializing in atmospheric chemistry

<u>Participant #2</u>

Please place an "X" or **bold** the appropriate response

1. Are you an employee of one of the following types of companies/ organizations/ institutions?

d. non-profit b. private
e. ***academia*** c. government
d. research/clinical e. other

2. Do you have any peer-reviewed publication on forecasting precipitation?

f. Yes b. ***No***

3. Do you have any peer-reviewed publication on weather forecasting?

g. Yes b. ***No***

4. Have you been nominated by your peers as an expert in forecasting precipitation?

h. Yes b. ***No***

5. Have you been nominated by your peers as an expert in weather forecasting?

i. Yes b. ***No***

6. Are you a member of professional organization *specifically* devoted to forecasting precipitation?
   j. Yes b. ***No***

7. Are you a member of professional organization devoted to forecasting weather?
   k. ***Yes*** b. No

8. Does your company/organization/institution specialize in forecasting precipitation?
   l. Yes b. ***No***

9. What is your Highest Level of Formal Education?
   m. Bachelor b. Master
   d. PhD

10. How many years of experience (beyond formal education), do you have in the field of weather forecasting?
   a. ***0 to <5 years*** b. 5 to <10 years
   c. >10 years

11. How many years of practical/field experience do you have in forecasting precipitation?
   a. <1 year b. ***1 to <5 years***
   c. 5 to <10 years d. >10 years
   e. none

12. Do you have any specialized certification in forecasting precipitation?
   n. Yes b. ***No***

13. Do you have any specialized certification in forecasting weather?
   o. Yes b. ***No***

14. List/Describe any other expertise qualifications:
   2nd place in the National Forecasting Contest Graduate Student Discussion 1985!

Participant #3

Please place an "X" or **bold** the appropriate response

1. Are you an employee of one of the following types of companies/ organizations/ institutions?

p. non-profit b. private
q. ***academia*** c. government
d. research/clinical e. other

2. Do you have any peer-reviewed publication on forecasting precipitation?

r. Yes b. ***No***

3. Do you have any peer-reviewed publication on weather forecasting?

s. Yes b. ***No***

4. Have you been nominated by your peers as an expert in forecasting precipitation?

t. Yes b. ***No***

5. Have you been nominated by your peers as an expert in weather forecasting?

u. Yes b. ***No***

6. Are you a member of professional organization *specifically* devoted to forecasting precipitation?

v. Yes b. ***No***

7. Are you a member of professional organization devoted to forecasting weather?

w. Yes b. ***No***

8. Does your company/organization/institution specialize in forecasting precipitation?

x. ***Yes*** b. No

9. What is your Highest Level of Formal Education?

y. Bachelor b. Master
e. PhD

10. How many years of experience (beyond formal education), do you have in the field of weather forecasting?

a. ***0 to <5 years*** b. 5 to <10 years
c. >10 years

11. How many years of practical/field experience do you have in forecasting precipitation?

a. ***<1 year*** b. 1 to <5 years
c. 5 to <10 years d. >10 years
e. none

12. Do you have any specialized certification in forecasting precipitation?

a. Yes b. ***No***

13. Do you have any specialized certification in forecasting weather?

b. Yes	b. ***No***

14. List/Describe any other expertise qualifications:

I have taught introductory Meteorology for two years. Most of my experience in weather forecasting stems from a basic interest in the weather, though I have only rarely issued forecasts, they were for a field experiment, and not posted for anyone else's use. In making those forecasts, I relied heavily on the work of others, as I did not have any direct access to models.

Participant #4

Please place an "X" or **bold** the appropriate response

1. Are you an employee of one of the following types of companies/ organizations/ institutions?

z. non-profit	b. private
aa. ***academia***	c. government
d. research/clinical	e. other

2. Do you have any peer-reviewed publication on forecasting precipitation?

bb. Yes	b. ***No***

3. Do you have any peer-reviewed publication on weather forecasting?

cc. Yes	b. ***No***

4. Have you been nominated by your peers as an expert in forecasting precipitation?

dd. Yes	b. ***No***

5. Have you been nominated by your peers as an expert in weather forecasting?

ee. ***Yes***	b. No

6. Are you a member of professional organization *specifically* devoted to forecasting precipitation?

ff. Yes	b. ***No***

7. Are you a member of professional organization devoted to forecasting weather?

gg. ***Yes***	b. No

8. Does your company/organization/institution specialize in forecasting precipitation?

hh. Yes b. ***No***

9. What is your Highest Level of Formal Education?
   ii. Bachelor b. ***Master***
   f. PhD

10. How many years of experience (beyond formal education), do you have in the field of weather forecasting?
   a. 0 to <5 years b. 5 to <10 years
   c. ***>10 years***

11. How many years of practical/field experience do you have in forecasting precipitation?
   a. ***<1 year*** b. 1 to <5 years
   c. 5 to <10 years d. ***>10 years***
   e. none

12. Do you have any specialized certification in forecasting precipitation?
   jj. ***Yes*** b. ***No***

13. Do you have any specialized certification in forecasting weather?
   kk. Yes b. ***No***

14. List/Describe any other expertise qualifications:

While in the US Navy I worked as a flight duty forecaster from 1989-1992.

APPENDIX E-1

SURVEY INSTRUMENT:

EXPERT JUDGMENT IN ADULT WEIGHT MANAGEMENT

EXPERT ATTRIBUTES DATA SHEET

Please place an "X" or **bold** the appropriate response

1. Are you an employee of one of the following types of companies/ organizations/ institutions?
   - o. non-profit
   - b. private
   - p. academia
   - c. government
   - d. research/clinical
   - e. other

2. Do you have any peer-reviewed publication on adult weight management?
   - q. Yes
   - b. No

3. Do you have any peer-reviewed publication in the field of Dietetics?
   - r. Yes
   - b. No

4. Have you been nominated/recognized by your peers as an expert in adult weight management?
   - s. Yes
   - b. No

5. Have you been nominated/recognized by your peers as an expert in the field of Dietetics?
   - t. Yes
   - b. No

6. Are you a member of any professional organization specifically devoted to weight management?
   - u. Yes
   - b. No

7. Are you a member of a professional organization devoted to dietetics?
   - v. Yes
   - b. No

8. Does your company/organization/institution specialize in adult weight management?
   - w. Yes
   - b. No

9. What is your highest level of formal education?
   - x. Bachelor's
   - b. Master's

c. PhD

10. How many years of experience (beyond formal education), do you have in the field of Dietetics?

a. 0 to <5 years   b. 5 to <10 years

c. >10 years

11. How many years of practical/field experience do you have in Adult Weight Management?

a. <1 year   b. 1 to <5 years

c. 5 to <10 years   d. >10 years

e. none

12. Do you have any specialized training or certification in Adult Weight Management?

y. Yes   b. No

List/Describe any other expertise that you may have in the field:

APPENDIX E-2

SURVEY INSTRUMENT:

<u>EXPERT JUDGMENT IN ADULT WEIGHT MANAGEMENT</u>

SAMPLE QUESTION SHEET

Please place an "X" or circle the correct response in questions 1-7 and 12.

1. A desirable weekly rate of weight loss in the overweight/obese adult is:
   a) 0.5-1.0 lbs   c) 2.0-3.0 lbs
   b) 1.0-2.0 lbs   d) 3.0-4.5 lbs

2. An approximate six months weight loss goal for George who is 250 lbs is:

   a) 10.5 lbs   c) 35 lbs
   b) 25 lbs   d) 42 lbs

3. In assessing risks for type 2 diabetes, hypertension and cardiovascular diseases, the predictive value of waist measurement of $\geq$ 40 inches in men is best at BMI of:
   a) 18.5   c) 25.0-29.9
   b) 18.5-24.9   d) 25.0-34.9

4. In the normal or overweight individual, which of the following is the best predictor of disease risk?
   a) BMI   c) waist circumference
   b) Body weight   d) mid arm circumference

5. Which of the following is not considered a risk factor for CVD?
   a) smoking   b) hypertension
   c) age $\geq$ 45 years for men   d) high levels of high density lipoprotein (HDL)

6. In treating Gill for obesity, the dietitian recommended a very low calorie diet. Gill would be consuming a diet that provides less than ___________ Kcal daily.

7. In order to achieve a desirable rate of weight loss, dietary calorie deficits should range from______ to _________daily.
   a) 300-450 Kcal   c) 500- 700 Kcal
   b) 450-550 Kcal   d) 500-1000 Kcal

8. 8. Which of the following is approved for long term use in the treatment of obesity?

a) sibutramine c) amphetamine

b) benzphetamine d) phentermine

For items 9-12, indicate "T" true or "F" false.

9. Obese adults are more prone to developing ketosis than children______

10. Zyprexa, risperdal and seroquel are medications noted for causing weight loss in the overweight/obese individual.______________.

11. In counseling clients it is best to ask "why" questions as these questions elicit "I don't know" responses or defensiveness.____________

12. In building rapport with clients, it is advisable to sympathize than to empathize with them. _________

SURVEY INSTRUMENT

EXPERT_A

EXPERT JUDGMENT IN ADULT WEIGHT MANAGEMENT

Expert Background

Please place an "X" by, or **"Bold"** the appropriate response

2. Are you an employee of one of the following types of companies/ organizations/ institutions?

a. non-profit	b. private
b. X academia	c. government
d. research/clinical	e. other

3. Do you have any peer-reviewed publication on adult weight management?

a. Yes	b.X No

4. Do you have any peer-reviewed publication in the field of Dietetics?

a. Yes	b.X No

5. Have you been nominated/recognized by your peers as an expert in adult weight management?

a. Yes	b.X No

6. Have you been nominated/recognized by your peers as an expert in the field of Dietetics?

a. Yes	b.X No

7. Are you a member of any professional organization specifically devoted to weight management?

a. Yes	b.X No

8. Are you a member of a professional organization devoted to dietetics?

a. XYes	b. No

9. Does your company/organization/institution specialize in adult weight management?

a. Yes	b. XNo

10. What is your highest level of formal education?

a. Bachelor's	b. Master's
g. XPhD	

11. How many years of experience (beyond formal education), do you have in the field of Dietetics?

a. 0 to <5 years b. 5 to <10 years
c. X>10 years

11. How many years of practical/field experience do you have in Adult Weight Management?
    a. <1 year b. 1 to <5 years
    c. 5 to <10 years d. >10 years
    e. none

12. Do you have any specialized training or certification in Adult Weight Management?
    a. Yes b.X No

List/Describe any other expertise that you may have in the field:
Several years of experience as a clinical dietitian who worked with obese individuals on a daily basis.

QUESTIONNAIRE

Please place an "X" by, or **"Bold"** the correct response in questions 1-7 and 12.

1. A desirable weekly rate of weight loss in the overweight/obese adult is:
    c) 0.5-1.0 lbs c) 2.0-3.0 lbs
    d) X 1.0-2.0 lbs d) 3.0-4.5 lbs

2. An approximate six months weight loss goal for George who is 250 lbs is:

    c) 10.5 lbs c) 35 lbs
    d) X 25 lbs d) 42 lbs

3. In assessing risks for type 2 diabetes, hypertension and cardiovascular diseases, the predictive value of waist measurement of $\geq$ 40 inches in men is best at BMI of:
    c) 18.5 c) 25.0-29.9
    d) 18.5-24.9 Xd) 25.0-34.9

4. In the normal or overweight individual, which of the following is the best predictor of disease risk?

a) BMI
b) Body weight
Xc) waist circumference
d) mid arm circumference

5. Which of the following is not considered a risk factor for CVD?

a) smoking
b) hypertension
c) age $\geq$ 45 years for men
Xd) high levels of high density lipoprotein (HDL)

6. In treating Gill for obesity, the dietitian recommended a very low calorie diet. Gill would be consuming a diet that provides less than _____400-800______ Kcal daily.

7. In order to achieve a desirable rate of weight loss, dietary calorie deficits should range from______ to _________daily.

c) 300-450 Kcal
d) 450-550 Kcal
Xc) 500- 700 Kcal
d) 500-1000 Kcal

8. Which of the following is approved for long term use in the treatment of obesity?

a) Xsibutramine
b) benzphetamine
c) amphetamine
d) phentermine

For items 9-12, indicate "T" true or "F" false.

9. Obese adults are more prone to developing ketosis than children___F___

10. Zyprexa, risperdal and seroquel are medications noted for causing weight loss in the overweight/obese individual._______F_______.

12. In counseling clients it is best to ask "why" questions as these questions elicit "I don't know" responses or defensiveness._____F_______

13. In building rapport with clients, it is advisable to sympathize than to empathize with them. _____F____

SURVEY INSTRUMENT

EXPERT_B

EXPERT JUDGMENT IN ADULT WEIGHT MANAGEMENT

Expert Background

Please place an "X" by, or **"Bold"** the appropriate response

1. Are you an employee of one of the following types of companies/ organizations/ institutions?

a. non-profit	b. private
b. **academia**	c. government
d. research/clinical	e. other

2. Do you have any peer-reviewed publication on adult weight management?

c. Yes	b. **No**

3. Do you have any peer-reviewed publication in the field of Dietetics?

d. **Yes**	b. No

4. Have you been nominated/recognized by your peers as an expert in adult weight management?

e. Yes	b. **No**

5. Have you been nominated/recognized by your peers as an expert in the field of Dietetics?

f. **Yes**	b. No

6. Are you a member of any professional organization specifically devoted to weight management?

g. Yes	b. **No**

7. Are you a member of a professional organization devoted to dietetics?

h. **Yes**	b. No

8. Does your company/organization/institution specialize in adult weight management?

i. Yes	b. **No**

9. What is your highest level of formal education?

j. Bachelor's	b. **Master's**
h. PhD	

10. How many years of experience (beyond formal education), do you have in the field of Dietetics?

a. 0 to <5 years b. 5 to <10 years
c. **>10 years**

11. How many years of practical/field experience do you have in Adult Weight Management?
a. <1 year b. 1 to <5 years
c. 5 to <10 years d. >10 years
e. **none**

12. Do you have any specialized training or certification in Adult Weight Management?
k. Yes b. **No**

List/Describe any other expertise that you may have in the field:

<u>QUESTIONNAIRE</u>

Please place an "X" by, or **"Bold"** the correct response in questions 1-7 and 12.

1. A desirable weekly rate of weight loss in the overweight/obese adult is:
e) 0.5-1.0 lbs c) 2.0-3.0 lbs
f) **1.0-2.0 lbs** d) 3.0-4.5 lbs

2. An approximate six months weight loss goal for George who is 250 lbs is:

e) 10.5 lbs **c) 35 lbs**
f) 25 lbs d) 42 lbs

3. In assessing risks for type 2 diabetes, hypertension and cardiovascular diseases, the predictive value of waist measurement of $\geq$ 40 inches in men is best at BMI of:
e) 18.5 c) 25.0-29.9
f) 18.5-24.9 **d) 25.0-34.9**

4. In the normal or overweight individual, which of the following is the best predictor of disease risk?
   a) BMI c) **waist circumference**
   b) Body weight d) mid arm circumference

5. Which of the following is not considered a risk factor for CVD?
   a) smoking b) hypertension
   c) age $\geq$ 45 years for men d) **high levels of high density lipoprotein (HDL)**

6. In treating Gill for obesity, the dietitian recommended a very low calorie diet. Gill would be consuming a diet that provides less than ______**1200__**___ Kcal daily.

7. In order to achieve a desirable rate of weight loss, dietary calorie deficits should range from______ to _________daily.
   e) 300-450 Kcal c) 500- 700 Kcal
   f) 450-550 Kcal **d) 500-1000 Kcal**

8. 8. Which of the following is approved for long term use in the treatment of obesity?
   a) **sibutramine** c) amphetamine
   b) benzphetamine d) phentermine

For items 9-12, indicate "T" true or "F" false.

9. Obese adults are more prone to developing ketosis than children____T__

10. Zyprexa, risperdal and seroquel are medications noted for causing weight loss in the overweight/obese individual._______F_______.

11. In counseling clients it is best to ask "why" questions as these questions elicit "I don't know" responses or defensiveness.________T____

12. In building rapport with clients, it is advisable to sympathize than to empathize with them. ______F___

SURVEY INSTRUMENT

EXPERT_C

EXPERT JUDGMENT IN ADULT WEIGHT MANAGEMENT

Expert Background

Please place an "X" by, or **"Bold"** the appropriate response

14. Are you an employee of one of the following types of companies/ organizations/ institutions?
    a. non-profit b. private
    b. **academia** c. government
    d. research/clinical e. other

15. Do you have any peer-reviewed publication on adult weight management?
    a. Yes b. **No**

16. Do you have any peer-reviewed publication in the field of Dietetics?
    a. Yes b. **No**

17. Have you been nominated/recognized by your peers as an expert in adult weight management?
    a. Yes b. **No**

18. Have you been nominated/recognized by your peers as an expert in the field of Dietetics?
    a. **Yes** b. No

19. Are you a member of any professional organization specifically devoted to weight management?
    a. Yes b. **No**

20. Are you a member of a professional organization devoted to dietetics?
    a. **Yes** b. No

21. Does your company/organization/institution specialize in adult weight management?
    a. Yes b. **No**

22. What is your highest level of formal education?
    a. Bachelor's b. Master's
    i. PhD

23. How many years of experience (beyond formal education), do you have in the field of Dietetics?

a. 0 to <5 years b. 5 to <10 years

c. **>10 years**

11. How many years of practical/field experience do you have in Adult Weight Management?

a. <1 year b. 1 to <5 years

c. 5 to <10 years d. >10 years

e. **none**

12. Do you have any specialized training or certification in Adult Weight Management?

a. Yes b. **No**

List/Describe any other expertise that you may have in the field:

QUESTIONNAIRE

Please place an "X" by, or **"Bold"** the correct response in questions 1-7 and 12.

1. A desirable weekly rate of weight loss in the overweight/obese adult is:

g) 0.5-1.0 lbs c) 2.0-3.0 lbs

h) **1.0-2.0 lbs** d) 3.0-4.5 lbs

2. An approximate six months weight loss goal for George who is 250 lbs is:

g) 10.5 lbs c) 35 lbs

h) **25 lbs** d) 42 lbs

3. In assessing risks for type 2 diabetes, hypertension and cardiovascular diseases, the predictive value of waist measurement of $\geq$ 40 inches in men is best at BMI of:

g) 18.5 c) 25.0-29.9

h) 18.5-24.9 d) **25.0-34.9**

4. In the normal or overweight individual, which of the following is the best predictor of disease risk?

a) BMI
b) Body weight
c) **waist circumference**
d) mid arm circumference

5. Which of the following is not considered a risk factor for CVD?

a) smoking
b) hypertension
c) age ≥ 45 years for men
d) **high levels of high density lipoprotein (HDL)**

6. In treating Gill for obesity, the dietitian recommended a very low calorie diet. Gill would be consuming a diet that provides less than ___1200________ Kcal daily.

7. In order to achieve a desirable rate of weight loss, dietary calorie deficits should range from______ to _________daily.

g) **300-450 Kcal**
h) 450-550 Kcal
c) 500- 700 Kcal
d) 500-1000 Kcal

8. 8. Which of the following is approved for long term use in the treatment of obesity?

a) sibutramine
b) benzphetamine
c) **amphetamine**
d) phentermine

For items 9-12, indicate "T" true or "F" false.

9. Obese adults are more prone to developing ketosis than children__T____

10. Zyprexa, risperdal and seroquel are medications noted for causing weight loss in the overweight/obese individual.________F______.

24. In counseling clients it is best to ask "why" questions as these questions elicit "I don't know" responses or defensiveness.___T_________

25. In building rapport with clients, it is advisable to sympathize than to empathize with them. ___F______

SURVEY INSTRUMENT

EXPERT D

EXPERT JUDGMENT IN ADULT WEIGHT MANAGEMENT

Expert Background

Please place an "X" by, or **"Bold"** the appropriate response

26. Are you an employee of one of the following types of companies/ organizations/ institutions?
    - a. non-profit — b. private
    - b. academia — c. government X
    - d. research/clinical — e. other

27. Do you have any peer-reviewed publication on adult weight management?
    - a. Yes — b. No X

28. Do you have any peer-reviewed publication in the field of Dietetics?
    - a. Yes X — b. No

29. Have you been nominated/recognized by your peers as an expert in adult weight management?
    - a. Yes — b. No X

30. Have you been nominated/recognized by your peers as an expert in the field of Dietetics?
    - a. Yes X — b. No

31. Are you a member of any professional organization specifically devoted to weight management?
    - a. Yes — b. No X

32. Are you a member of a professional organization devoted to dietetics?
    - a. Yes X — b. No

33. Does your company/organization/institution specialize in adult weight management?
    - a. Yes — b. No

34. What is your highest level of formal education?

a. Bachelor's b. Master's
j. PhD X

35. How many years of experience (beyond formal education), do you have in the field of Dietetics?
a. 0 to <5 years b. 5 to <10 years
c. >10 years X

11. How many years of practical/field experience do you have in Adult Weight Management?
a. <1 year b. 1 to <5 years
c. 5 to <10 years d. >10 years
e. none X

12. Do you have any specialized training or certification in Adult Weight Management?
a. Yes X b. No

List/Describe any other expertise that you may have in the field:

Teaching and research expereince

QUESTIONNAIRE

Please place an "X" by, or **"Bold"** the correct response in questions 1-7 and 12.

1. A desirable weekly rate of weight loss in the overweight/obese adult is:
i) 0.5-1.0 lbs c) 2.0-3.0 lbs
j) 1.0-2.0 lbs X d) 3.0-4.5 lbs

2. An approximate six months weight loss goal for George who is 250 lbs is:

i) 10.5 lbs c) 35 lbs
j) 25 lbs d) 42 lbs X

3. In assessing risks for type 2 diabetes, hypertension and cardiovascular diseases, the predictive value of waist measurement of $\geq$ 40 inches in men is best at BMI of:

i) 18.5
j) 18.5-24.9 X
c) 25.0-29.9
d) 25.0-34.9

4. In the normal or overweight individual, which of the following is the best predictor of disease risk?

a) BMI
b) Body weight
c) waist circumference X
d) mid arm circumference

5. Which of the following is not considered a risk factor for CVD?

a) smoking
b) hypertension
c) age $\geq$ 45 years for men
d) high levels of high density lipoprotein (HDL) X

6. In treating Gill for obesity, the dietitian recommended a very low calorie diet. Gill would be consuming a diet that provides less than ___________ Kcal daily.

7. In order to achieve a desirable rate of weight loss, dietary calorie deficits should range from______ to _________daily.

i) 300-450 Kcal
j) 450-550 Kcal
c) 500- 700 Kcal
d) 500-1000 Kcal X

8. 8. Which of the following is approved for long term use in the treatment of obesity?

a) sibutramine
b) benzphetamine
c) amphetamine
d) phentermine

For items 9-12, indicate "T" true or "F" false.

9. Obese adults are more prone to developing ketosis than children_____F_

10. Zyprexa, risperdal and seroquel are medications noted for causing weight loss in the overweight/obese individual.__________T____.

11. In counseling clients it is best to ask "why" questions as these questions elicit "I don't know" responses or defensiveness._______F_____

12. In building rapport with clients, it is advisable to sympathize than to empathize with them. __F_______

APPENDIX-F

PARTIAL LISTING OF SOURCES

Database	Articles Generated by Search
WorldCat	600
Agricola	800
DOE’s Information Bridge	200
Civil Engineering	19
Energy Citations	223
Waste Management research abstracts	30
PubMed	1146
Medline	0
Dissertation Abstracts	191
Total	3209