Understanding and Strengthening Exposure Judgments Using Bayesian Integrated Exposure Assessment Strategies

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Perry William Logan

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Gurumurthy Ramachandram, Ph.D.

December, 2010

Acknowledgements

I would like to thank John Mulhausen, Gurumurthy Ramachandran, Paul Hewett and Bruce Alexander for all of their technical guidance, friendly debate and passion for enhancing the core science of environmental and occupational hygiene. I would like to also thank Carleen Andert for all of her amazing support, patience and love through so many long weekends and late nights supporting this work. My sincere gratitude goes out to all of my wonderful collogues and friends who continue engaging this work to strengthen the science in exposure assessment.

Dedication

This work is dedicated to all of the exposure assessment professionals who work hard making scientific contributions to the environmental and occupational health profession.

Table of Contents

List of Tables

List of Figures

CHAPTER I. Introduction – Understanding Exposure Judgment Accuracy

SPECIFIC AIMS

Accurate exposure assessments by occupational hygienists are one of the most critical elements to ensuring that hazardous exposures are properly controlled and workers are adequately protected. The availability and accuracy of exposure assessments can determine whether or not resources applied to engineering and administrative controls, medical surveillance, personal protective equipment and other programs effectively protect workers. Exposure assessments are also used to provide exposure input for epidemiological studies investigating the prevalence of disease related to occupational and environmental exposure of disease agents within a population. In most cases, exposure assessments made for a given job or worker are subjective judgments made by occupational hygienists with limited sampling data. These judgments are assumed to be accurate and unbiased; however they are rarely studied and validated. A Bayesian approach (Hewett et al, 2006) to the American Industrial Hygiene Association (AIHA) Exposure Assessment and Management Strategy (Mulhausen and Damiano 1998; Ignacio and Bullock 2006) shows great potential to enhance the accuracy and efficiency of the exposure assessment process. A great benefit of the approach is the opportunity for a transparent method to integrate, test accuracy, identify judgment bias, leverage all relevant exposure information and validate exposure judgments made with limited sampling data.

The AIHA process of making exposure judgments integrates qualitative information based on experience and observation as well as quantitative exposure monitoring data that may exist for a given operation. This process is iterative and allows for updating of judgments as new information or data becomes available. Hygienists bring together available sources of qualitative and quantitative information to arrive at an integrated exposure judgment using some type of mental model or reasoning developed from their education and experience. This process is inherently heuristic in nature and in some cases may contain cognitive biases because of the unstructured methods used to integrate the various sources of information. This identifies a need for a more formal and structured Bayesian approach using the AIHA strategy to convert qualitative and quantitative information into a probability framework creating a transparent mathematical combination of these various sources of information and data. A promising aspect of this Bayesian approach is that using probabilities and the mathematical techniques appear to intuitively represent the process of making an integrated judgment using mental heuristics or “professional judgment” and therefore should provide valuable feedback. Since this Bayesian approach matches the AIHA strategy, it can be easily adapted or integrated into existing processes and systems that utilize a method based on the well-known AIHA strategy (Hewett et al, 2006).

Qualitative exposure information and data can come from many sources including past experience, surrogate sampling data, specific training, physico-chemical models, empirical models and others. The efficiency of this approach can increase when this qualitative information is used in conjunction with exposure monitoring data to arrive at an accurate final judgment with fewer sampling data. In order to obtain maximum benefit from this approach, qualitative or professional judgments made with no or limited exposure monitoring data must be reasonably accurate and factors affecting their accuracy must be well understood. With a better understanding of factors that affect professional judgment accuracy and bias, specific training can be developed to improve efficiency and better leverage professional judgment. In addition, this understanding may also lead to better design of tools, programs and systematic interventions that can control biases, increase efficiency and effectiveness of exposure assessment and management programs.

This research was focused to evaluate the following:

- accuracy of categorical exposure judgments made by occupational hygienists,
- factors or "determinants" of education and experience that affect judgment accuracy,
- performance of a Bayesian integrated and other various exposure assessment strategies, and
- methods to increasing exposure judgment accuracy.

To study exposure judgment accuracy, desktop elicitations and a probabilistic categorical judgment framework were used to facilitate evaluation of accuracy, bias and relationships with determinants. An accurate exposure assessment is defined as one whereby a hygienist correctly identifies an exposure category where the 95th percentile of the exposure distribution falls. Although not used in this study, the approach outlined in this research proposal can also facilitate using the mean, median, geometric mean or other upper tail percentiles as the decision statistic. A central tendency decision statistic would be more appropriate for applications in retrospective exposure judgments of chronic disease agents for epidemiological studies. However, an upper percentile decision statistic is more appropriate for ensuring that most all daily exposures in a Similar Exposure Group (SEG) fall below a given OEL.

Exposure categories can be defined as fractional and multiple ranges of a given occupational exposure limit (OEL) so that the exposure judgment can be defined as the exposure category where the decision statistic most likely falls. The four categories selected from the AIHA strategy used in this study are; <10%, 10-50%, 50-100% and >100% of the OEL (Table I.I). These categories are used by the hygienist to make an exposure judgment by interpreting exposure information and relevant monitoring data. The exposure judgment is translated into the probability of the decision statistic falling in each of the 4 categories. In some cases in this study, it is valuable to restrict the use of statistical tools for interpreting data to better study upper tail judgments. This method can illustrate how well a hygienist may perform exposure judgments without the use of statistical tools when data is available. Since judging the decision statistic of

lognormal datasets is central to performing prospective exposure assessments, a Data Interpretation Test (DIT) is devised to help test and calibrate a hygienist's ability to estimate 95th percentiles of distributions based on small, simulated data sets. The DIT can also provide a data interpretation accuracy score that will be used as a determinant in the study. In the future, the DIT may become a surrogate test of one's ability to make accurate exposure judgments with very limited sampling data.

It is hypothesized that if exposure judgments are reasonably accurate, they can be incorporated into a Bayesian decision-making procedure that synthesizes the subjective judgments with monitoring data to provide updated judgments that are more effective and efficient, i.e., correct judgments are arrived at in the least number of measurements (Hewett et al, 2006). This goal was addressed by studying accuracy and determinants of exposure judgment for several groups of occupational hygienists from different backgrounds in education, industry, organizations and experience with documenting chemical agent exposure judgments. The main questions for this research were:

1. Are exposure judgments made by occupational hygienists more accurate than random chance?
2. Is specific training effective in improving judgment accuracy?
3. What education and experience determinants influence the ability to make accurate exposure judgments?
4. Does a Bayesian integrated exposure assessment strategy perform better than other common strategies?

The following actions were taken to generate the exposure judgment data used in this study:

a) A group of job tasks common to different types of manufacturing processes, which range across the 4 different AIHA exposure categories, were selected. A data collection framework that was used to obtain probabilistic qualitative exposure judgments for specific tasks using the 4 AIHA exposure categories without the use of statistical tools. This framework was used to test the ability of occupational hygienists to interpret small sample data sets without statistical computing tools. This test known in this study as the Data Interpretation Test or "DIT" was used as a metric for an individual's data interpretation abilities in the absence of statistical tools.

b) Basic characterization information was provided by videos of each task being performed, along with written information describing the task, materials, equipment, exposure controls and other details that can be used to make exposure judgments. The occupational hygienists were asked to provide their exposure judgments by estimating the probability of the 95th percentile being in each of the 4 AIHA exposure categories. In order to define the actual exposure for each task or 'truth', a sufficient number of measurements of task exposures (8-10 measurements per task) were collected to estimate the exposure distribution. The probabilistic exposure distribution obtained solely from these measurements

will be assumed to be the 'truth' that determines the AIHA category where the population 95th percentile most likely falls.

c) Groups of occupational hygienists were assembled who have performed exposure judgments in various organizations with varying levels of seniority, manufacturing facility experience, exposure assessment and modeling experience, educational backgrounds and training. A process to collect background information from each hygienist regarding experience, education, certification, statistics and data interpretation skills was developed. Participants were trained on the purposes of this study and the procedures used to obtain exposure judgments and determinants. For some judgment elicitations, participants were arranged randomly into groups of 3 to 4 participants and elicit a group judgment for each task on each full dataset to allow for comparison of individual and group final judgments. Exposure judgments were elicited from participants for the tasks selected without sampling data and then successively when each sampling data point was revealed. A different DIT was administered before and after all task exposure judgment data were collected. Exposure judgments and DITs were collected before and after a data interpretation training was delivered to the group of participants.

d) Simulations were performed to illustrate how several exposure assessment strategies can properly detect acceptable and unacceptable exposures. Exposure populations with exceedance fractions of 0.01%, 0.1%, and 1% wereused as 'acceptable' and exceedance fractions of 10%, 20%, 30% and 50% were used as 'unacceptable'. Geometric standard deviation and number of samples were used as important variables in the simulations to compare the performance of each selected strategy.

BACKGROUND AND SIGNIFICANCE

Current Methods for Making Exposure Judgments and the Limitations

Occupational hygienists utilize skills that can be considered a mixture of science and art to anticipate, recognize, evaluate and control workplace hazards. An ultimate goal for any exposure assessment would be to accurately define the exposure distribution for a single person or group of people over the time frame of interest. Hypothetically, this would be quite easy if all individuals in a given study population wore accurate agent sensing devices that measured and stored their real-time exposure for a given agent or group of agents. Exposure assessment would then become a routine data retrieval exercise and virtually eliminate the need for qualitative exposure assessments. Until these futuristic sampling devices become available, qualitative exposure assessments will serve as a foundation for exposure assessment and management programs.

The ability to utilize all available information to accurately evaluate or judge exposure is at the heart of making accurate exposure assessments. As the process of making exposure assessments matures to become

more documented and efficient, the structure can be used to test judgment accuracy and provide a feedback mechanism that can be used to increase judgment accuracy, thereby improving the ability to better recognize and control hazards. A well-known strategy for exposure assessment and management can be found in the Second and Third Editions of AIHA's Exposure Assessment and Management Book (Mulhausen and Damiano 1998; Ignacio and Bullock 2006). The strategy recommends that all exposures be characterized qualitatively initially and, where necessary, refined quantitatively by collecting samples that estimate worker exposure. The strategy implies that professional judgment can be combined with sampling data to arrive at a more accurate and efficient decision, which lends to a good integration with Bayesian statistical methods. Exposure judgment efficiency can be defined as arriving at a correct decision with the least number of exposure monitoring data points. It has been suggested by Hewett et al (2006) that by incorporating a Bayesian statistical approach, the AIHA strategy can be further refined to create a more transparent process for understanding and incorporating professional judgment into the exposure assessment and management process. This method will utilize Bayesian approaches to use quantitative exposure judgments for validation of qualitative or semi-quantitative exposure judgments. This proposed study will utilize a probabilistic exposure assessment framework based on the AIHA model to analyze sampling data and exposure judgments made by occupational hygienists. In addition, several statistical methods will be used to identify determinants that impact accuracy in qualitative and semi-quantitative exposure judgments.

Many of the published strategies share the idea of initially classifying workers into similar exposure groups (SEGs) based on observation of the process. Well known strategies for classification of SEGs have been proposed and used (Corn and Esmen, 1979; Mulhausen and Damiano 1998). Occupational hygienists review the workforce, materials, exposure agents, tasks, equipment, exposure controls and identify exposure groups that will be assessed and possibly controlled depending on the final judgments. The exposure judgment for any SEG requires the selection of an Occupational Exposure Limit (OEL) and a judgment by the hygienist about where the decision statistic (e.g., the 95th percentile of the exposure distribution for the SEG) falls in relation to the OEL. The AIHA strategy illustrates a method for defining control categories or exposure ratings around the OEL for each exposure agent (e.g., Table I.I).

Table I.I: Exposure category rating scheme. A Similar Exposure Group or SEG is assigned an exposure rating by comparing the 95th percentile exposure ($X_{0.95}$) of the exposure distribution with the full-shift, TWA occupational exposure limit, OEL or STEL (Short Term Exposure Limit), to determine which category it most likely falls into. (Mulhausen and Damiano, 1998)

AIHA Exposure Rating	Proposed Control Zone Description	General Description	AIHA Recommended Statistical Interpretation
1	Highly-controlled (HC)	95th Percentile of exposures rarely exceeds 10% of the limit.	$X_{0.95} \leq 0.10$ OEL
2	Well-controlled (WC)	95th Percentile of exposures rarely exceeds 50% of the limit.	$0.10 \text{ OEL} < X_{0.95} \leq 0.5$ OEL
3	Controlled (C)	95th Percentile of exposures rarely exceeds the limit.	$0.5 \text{ OEL} < X_{0.95} \leq \text{OEL}$
4	Poorly controlled (PC)	95th Percentile of exposures exceeds the limit.	$\text{OEL} < X_{0.95}$

The occupational hygienist assigns an initial "exposure rating" which represents the category where the decision statistic most likely falls based on the information available at the time. In this proposal, the decision statistic selected is the 95th percentile of the population exposure distribution. Based on the certainty of the assessment and the hazards posed by the exposure, SEGs are then prioritized for follow-up quantitative studies (i.e., the collection of additional workplace exposure measurements), immediate control implementation or both. SEGs that merit a high initial exposure rating, low certainty in the exposure judgment, and involve high toxicity substances should be placed at the top of the priority list for follow-up actions. SEGs at the opposite end with low exposure ratings and high certainty should be lower in the priority list. "Certainty" of a judgment is qualitative and somewhat analogous to the width of a statistical confidence interval. An example is an exposure judgment with a high certainty, producing a narrower confidence interval and more probability falling into a single exposure category.

Most commonly, the hygienist uses a combination of professional judgment, personal experience with a given type of operation, review of exposures or monitoring data from similar operations, and/or exposure predictions developed using physical/chemical exposure modeling techniques to place SEGs into one of the control categories. Experience in making exposure judgments is likely to be related to accuracy and certainty in exposure judgments, probably much like other experiences in life. As an individual gains more experience performing a particular job or activity, the more confident they become about their abilities.

Therefore, it is important that feedback mechanisms for accuracy of judgments exist so that inaccuracy and bias can be reduced and not perpetuated to other exposure judgments. Certainty and accuracy should increase as a hygienist gains experience evaluating specific job-agent exposures with a reliable feedback mechanism.

The AIHA strategy recommends randomly sampling 6-10 events of a specific SEG and calculating an upper tail decision statistic such as the 95th percentile with an upper confidence limit (e.g. 90th or 95th). This information can be used to select the AIHA exposure control category where the decision statistic most likely falls. In many cases with limited monitoring data, the decision statistic point estimate and upper confidence limit fall into different categories and monitoring data is limited for an SEG, the AIHA strategy recommends that professional judgment be used to resolve the final judgment for the most likely exposure control category. Information that can be used to strengthen professional judgment includes past experience with a similar exposure scenario, utilizing surrogate data on similar SEGs, physical-chemical modeling, and other available empirically based exposure models (Jayjock 1997; Keil 2000; Nicas 2003;Mulhausen and Damiano 1998; Ignacio and Bullock 2006). Limitations with this approach include:

- Difficulty in calculating an upper tail decision statistic with upper confidence limit and selecting a most likely exposure control category with small data sets (<6 measurements).
- Professional judgment and monitoring data are intuitively combined by the hygienist, which can introduce inconsistent, inaccurate and biased judgments.
- Understanding the various inputs used to determine an exposure control category may be difficult because of the intuitive nature used to combine professional judgment and monitoring data.

Bayesian Decision Analysis

Unlike conventional statistical methods applied to occupational exposure data, Bayesian statistical techniques are designed to explicitly utilize not only monitoring data but also professional judgment and other sources of information. The “Prior” probability distribution is a chart of the decision probabilities illustrated below in Figures 1.1(a) and (b) that were developed using professional judgment. The other input in the Bayesian analysis is exposure monitoring data used to develop the Bayesian “Likelihood” which is displayed in the same format as the “Prior”. The Bayesian calculations use both the “Prior” and “Likelihood” to generate an integrated acceptability judgment or Bayesian “Posterior”, is illustrated in Figure 1.2. This approach leverages information and data together into a final judgment that helps provide efficiency and transparency to the ‘art’ of making exposure judgments. It is hypothesized that the ultimate result will be more accurate exposure judgments made with fewer exposure monitoring data.

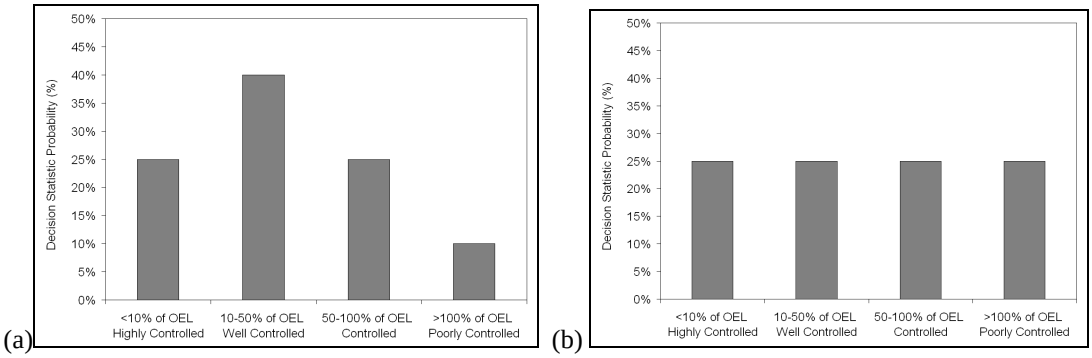

Figure 1.1(a). This example shows an exposure judgment for a given SEG using exposure categories. The bar chart shows the probability that the decision statistic (the 95th percentile of the exposure distribution) is most likely between 10% and 50% of the Occupational Exposure Limit (OEL). Since the four categories cover all possible ranges for the Decision Statistic, the probabilities from all four categories must add to 100%. Figure 1.1(b). Uninformed professional judgment prior indicating that is utilized when there is no additional information or data to leverage with the monitoring data likelihood.

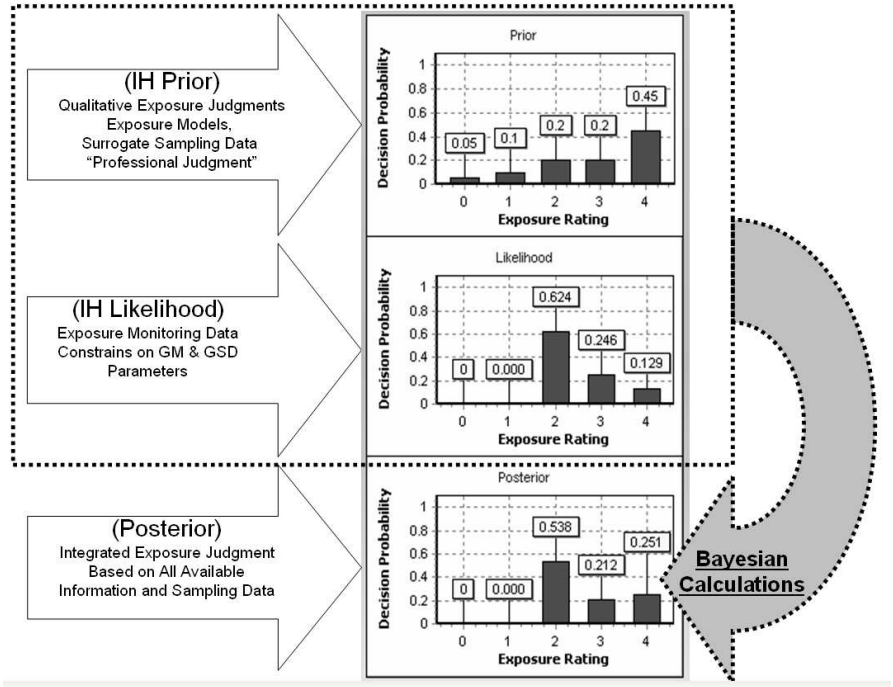

Figure 1.2. Pictorial representations of a Bayesian approach to the exposure judgment process.

The process of utilizing available information and past experience to arrive at a probabilistic professional judgment "Prior" gives us a framework to more rigorously study factors that affect the accuracy and bias of qualitative exposure judgments. In situations where there is an abundance of exposure monitoring data or

where there is a very high lack of certainty in professional judgment, the final decision can be driven solely by the monitoring data. SEGs with a low certainty in professional judgment can be analyzed using an *un-informed or flat prior*, which has equal probability in each exposure category (Figure 1.1(b)). With situations having no exposure monitoring data, the final decision or "Posterior" is driven solely by the "Prior" qualitative judgment.

Professional judgment accuracy and bias can be directly studied by carefully collecting adequate exposure monitoring data to define 'truth' and comparing judgments made without the use of the monitoring data and statistical tools. This framework is fundamental to this study and provides a systematic approach to use quantitative exposure judgments for testing the accuracy and bias of qualitative or semi-quantitative exposure judgments.

Several important advantages of incorporating a Bayesian approach into the AIHA Exposure Assessment framework include:

1. Physical-chemical models and professional judgment can be directly incorporated with exposure monitoring data to leverage all available information to arrive at decisions more quickly and transparently.
2. Monitoring data can be used in conjunction with objective information on the SEG's geometric standard deviation (GSD) and geometric mean (GM) to arrive at a Bayesian Likelihood probability chart. This constraint on the parameters allow for hygienists to specify the most likely exposure control category even with low sample sizes. (Hewett et al 2006)
3. Common biases identified from studies of professional judgment in other fields may also be significant in occupational exposure assessments which include: base-rate errors, belief in the law of small numbers, false anchoring, availability, and overconfidence. (Kahneman and Tversky, 1982; Gilovich et al, 2002). This new framework creates a transparent and systematic method for studying judgment accuracy and bias by directly comparing probabilities in the prior and likelihood charts.
4. Systematically utilizing likelihood outputs from actual monitoring surveys can provide occupational hygienists an excellent feedback mechanism to help calibrate exposure judgments, identify determinants affecting judgment accuracy and design specific strategies to increase professional judgment accuracy. (Hewett et al, 2006)

Probabilistic Exposure Judgments – Quantifying Accuracy and Certainty

One can think of an exposure judgment as being defined in two parts: exposure rating and certainty (Mulhausen and Damiano, 1998). If a hygienist investigates a SEG that has limited information regarding

the amounts a specific chemical used or the effectiveness of exposure controls, the hygienist will still make an exposure judgment but may have a low level of certainty about the judgment. Regardless of the judgment certainty for the SEG, the hygienist always makes an exposure judgment to determine if exposure controls or further information gathering are needed. As more information and sampling data becomes available, the certainty regarding the exposure judgment will increase to a point where no additional information is needed.

An approach proposed by (Hewett et al, 2006) helps account for both the exposure judgment and certainty by using a probabilistic framework to allocate probabilities for the Decision Statistic falling into each AIHA exposure category. The approach leads hygienists to qualitatively estimate the category where the decision statistic for the exposure profile is most likely located using probabilities. This approach does not require that a precise value of the decision statistic be known in order to make a decision and therefore makes the strategy more efficient because selecting categories requires less statistical precision. Using exposure control categories or bands is more aligned with common approaches since it is much easier to communicate and manage.

Figure 1.1(a) illustrates this method using the AIHA exposure control categories for an SEG that has approximately 40% probability of the decision statistic falling between 10% and 50% of the OEL. In this case the hygienist has concluded after reviewing all of the applicable information and data that the 95th percentile has the highest probability of being in the category between 10% and 50% of the OEL. However, there is still a significant probability that the decision statistic is greater than the OEL which could likely require exposure controls or at least further data collection and possibly exposure monitoring. This example shows how a lack of data results in a low level of certainty expressed as relatively similar probabilities in all 4 categories.

Let us consider the same hygienist and a different SEG where multiple samples have been collected which were all below 10% of the OEL. The hygienist reviewed work practices in addition to an inspection of the ventilation controls indicating exposures are most likely highly controlled. Figure 1.3 illustrates this example of the hygienist's updated exposure judgment based on the new information and data. This is an example where the hygienist has a high degree of certainty regarding the judgment.

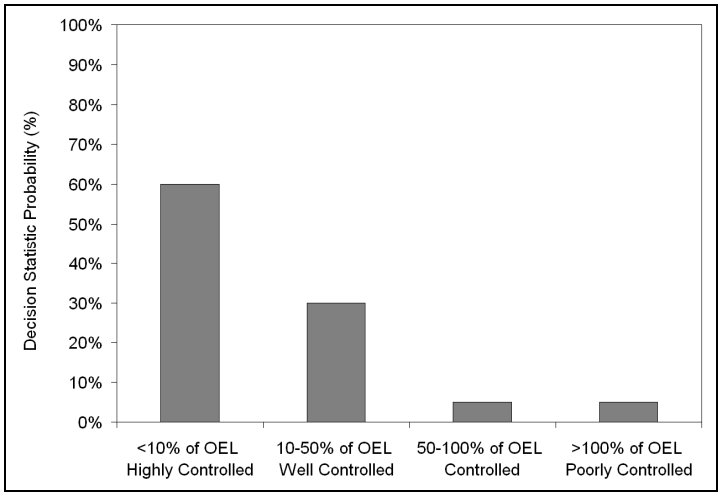

Figure 1.3. Example exposure judgment probability chart for a given SEG where a hygienist has more information regarding well designed exposure controls and all surrogate sampling data well below the OEL indicating that exposures most likely fall in the category below 10% of the OEL.

However, a difficulty with this or any exposure management strategy comes when hygienists are required to make exposure judgments with very limited information and data. Any strategy where occupational hygienists make exposure judgments without adequate information or data has the potential to introduce inaccuracy and bias which could leave people unprotected. The ambiguous process of making exposure judgments with inadequate information has sometimes been referred to as the art of professional judgment. The use of the Bayesian framework helps to introduce standardization, mathematical rigor and a structure for validation to make the process more transparent and effective.

Heuristics and Biases in Judgment

The current literature in psychology relating to human judgment with uncertainty proposes that a limited number of simplifying heuristics are used to efficiently arrive at a judgment using available information. These heuristics or mental processes do not typically utilize all of the available information and data in a formal algorithmic process but use more of "fast and frugal" rules of thumb to arrive at a judgment. (Kahneman and Tversky, 1982; Gilovich, 2002; Gigerenzer, 1999) Heuristics can be thought of as simple, efficient rules, hard-coded by evolutionary processes which have been proposed to explain how people make decisions, come to judgments, and solve problems, typically when facing complex problems or incomplete information. The concept of heuristics became very popular after the 2002 Nobel Prize in Economics was awarded to Kahneman and Tversky for their work in Prospect Theory which defined several heuristics in economics related to purchasing decisions (Kahneman and Tversky, 1979). More

recently the theory of heuristics has taken on a more encompassing scope which includes both intuitive and reflective or conscious decision making systems (Gigerenzer, 1999; Gilovich, 2002)

Studies on various heuristics and their common biases cover a wide range of fields including psychology, economics, medicine and engineering. They indicate that humans use simple heuristics rather than extensive algorithms when making judgments that each has biases associated with them because of their simplistic nature (Kahneman and Tversky, 1982; Griffin, 1992; Gigerenzer, 1999). These simple and efficient cognitive processes in some cases may be biased because of flaws in their simple algorithms (Kahneman and Tversky, 1982). The three heuristics proposed (Kahneman and Tversky, 1982) continue to be studied in a variety of fields include: (1) anchoring and adjustment, (2) availability and (3) representativeness. There are different biases that can be a result of the frugal rules used in the heuristic or from the available information and possibly a combination of both. The amount of bias related to each of these heuristics is usually related to the certainty of the judgment and availability of supporting information and data (Kahneman and Tversky, 1982). Kahneman and Tversky's experience teaching statistics and their observations of predictions made in applied settings led them to investigate sources of error and bias. Early on they concluded that people fail to anticipate regression to the mean, fail to give adequate weight to sample size in assessing the import of evidence, fail to take full advantage of base rates when making predictions and possibly other unknown cognitive biases. Their three heuristics were offered as an explanation as to why some errors may occur in a variety of professions (Gilovich, 2002).

The anchoring and adjustment heuristic may be used when judgments are made using a single data point or piece of information to make a judgment. People will tend to use the single data point or anchor and make adjustments from it to arrive at the final judgment. An example is a person who is looking to buy a car using only the odometer reading and year to make a final judgment on a used car purchase without evaluating maintenance records, accident history and test drive. This purchase judgment would be efficient, requiring only 2 data points, however would likely be biased and inaccurate (Kahneman and Tversky, 1982). The human brain has a very fast and powerful memory which serves as the architecture for the availability heuristic. Judgments always require some utilization of memory processing and can be driven by what information is available and most salient in memory. It is normal and quite correct in many cases to assume commonality with things that appear similar. This is the basis for the representativeness heuristic which uses memory and cognitive features together to form judgments quickly and efficiently (Kahneman and Tversky, 1982; Griffin, 1992; Gigerenzer, 1999).

To illustrate these 3 heuristics using a hypothetical example for exposure assessment, consider a hygienist that has 4 years experience working in the chemical industry, and is asked to estimate the upper tail exposure distribution for workers handling a new material containing small amounts of methylene chloride on a new process using a special filter press. The hygienist may use the *availability* heuristic to recall from

memory all past exposure assessments on a similar filter press using a different material containing xylene. The hygienist may also utilize the *representativeness* heuristic and may pour over the basic characterization information to identify the similarity of this new material, chemicals and process with the recalled exposure assessments to determine if they likely match. In addition the hygienist may use the *anchoring and adjustment* heuristics to take information recalled from a representative process exposure assessment and adjust the upper tail judgment higher based on the higher vapor pressure of methylene chloride than xylene. The most likely case would be for the hygienist to utilize a combination of these and other heuristics to arrive at the final judgment.

Kahneman and Tversky's three popular heuristics first proposed are probably only a small subset of other heuristics that exist commonly. In addition, they are not each used exclusively but may be used together and with other unknown heuristics in different combinations (Griffin, 1992). Studies indicate that it is not possible to determine a priori which heuristic governs a given judgment even when a limited number of heuristics are investigated (Gilovich, 2002). Consider the example where people may use different heuristics when asked whether more people die from rattlesnake bits or from bee stings. If a person had recent heard a news report about snake bites in a given state, they may utilize the availability heuristic by using the relative availability of the two choices when making the judgment. Alternatively, a person may not have any instances come to mind and might consider the relative "dangerousness" of snake bites and bee stings which would be an example of the representativeness heuristic (Gilovich, 2002).

Some studies indicate that heuristics used in human judgment originate from different sources, some are hard coded through the evolutionary process and some can be obtained through conscious learning. A study of Chinese 4th, 5th, 6th graders and MBA students showed that the level of education strongly influenced the ability to solve Bayesian problems. The authors indicated that one likely reason for the finding was that educational training instilled heuristics that helped solve Bayesian problems. (Zhu, 2006) This study and others showed that people may be more successful making judgments using frequency formats rather than probability formats without additional training or education. It does appear that training and education helps to close the gaps in judgment effectiveness between frequency and probability formats (Gigerenzer 1999). Several authors have concluded that humans likely use a form of dual processing consisting of intuitive heuristics and rational or reason based heuristics. The intuitive heuristics may rely more on "feeling" or emotions while the rational heuristics use clearly cognitive rules of thumb or simple algorithms.

Several studies indicate that overconfidence is very common and propose several possible determinants of certainty (Griffin, 1992). They suggest that certainty of judgments is based both on the strength of evidence and the credibility or weight of the evidence. The authors propose that judgment overconfidence is determined by the strength and balance of arguments for and against the competing hypotheses, without

proper regard for the weight of the evidence such as sample size. The authors give many examples of overconfidence in judgment making with non-experts and also professionals across several fields. In addition, the authors present study results that show how overconfidence can be caused when the sample parameter values are high and the sample size is low. They also show in the same study that underconfidence was created when the sample parameter values were low and the sample size was high.

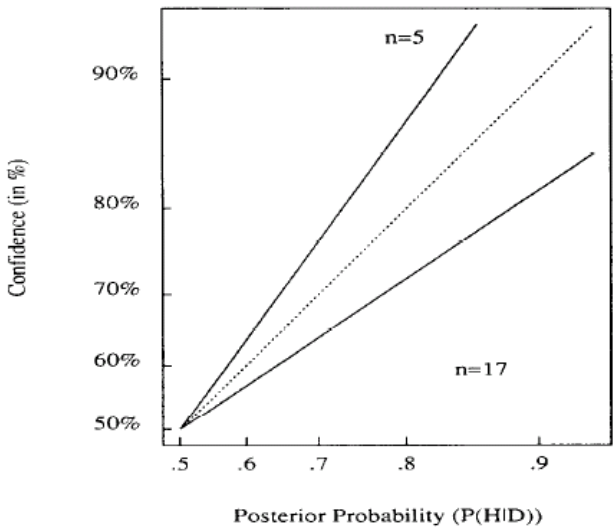

Figure 1.4 – From Griffin 1992, a graph illustrating the effect of sample size on judgment confidence.

An example illustrating overconfidence for occupational hygiene would be to consider a hygienist making an exposure assessment on a new process by utilizing only 1 exposure monitoring result at 10% of the OEL. If the hygienist judged the upper tail exposure to be less than 10% of the OEL with high certainty but without previous experience, adequate basic characterization and supporting exposure assessment information, it would be an example of an overconfident judgment. Overconfidence in exposure judgments is likely to come from not properly accounting for the uncertainty with limited sample data and experience with given exposure scenarios, not understanding the exposure population base rate, or other biases introduced from a variety of heuristics.

The Use of Professional Judgment in Occupational Exposure Assessment

As mentioned earlier and illustrated by the AIHA strategy, professional judgment is commonly used to assess exposures where monitoring data is limited or not yet available. The process of making an exposure judgment with no monitoring data has been periodically studied and published. Many have indicated that this ‘art’ of making exposure judgments sometimes referred to as ‘expert judgment’ is some combination of professional experiences, educational background and other unknown factors. (Kromhout et al., 1987;

Hawkins and Evans, 1989; Post et al., 1991; Cherrie and Schneider, 1999; Ramachandran and Vincent, 1999; Ramachandran, 2001).

Macaluso *et al*. (1993) studied a pool of 695 "department-job-year" exposure combinations made by five experts that provided assessments in one of seven exposure categories. The inter-rater agreement was considered good for the frequent exposure jobs however the low exposures had poor concordance among raters. The authors suggested that further study is needed to investigate the determinants of the judgments to evaluate potential sources of bias. Cock *et al*. (1996) studied exposure assessment related to potential dermal and respiratory exposure related to pesticide application in fruit farms. They utilized judgments from 15 experts who ranked the exposure for 14 different pesticide application tasks on fruit farms. The study used Cohen's kappa statistic and an intra-class correlation coefficient determined using ANOVA to measure inter-rater agreement. The exposure ratings made by the experts were found to correlate with their level of expertise making exposure assessments. The results of the study showed a statistically significant correlation between the exposure rankings and actual measured exposures.

Studies performed by Kromhout *et al*. (1987) looked at estimating task exposures made by occupational hygienists, supervisors, and workers at a number of plants. Each subject rated exposures across five plants for 97 tasks, into one of four categories ranging from no exposure to high exposure. Significant correlation between the ratings and the measured mean exposures was found which led the authors to recommend methods that combine subjective exposure estimates and actual measurements for better retrospective exposure estimates. This recommendation is similar to others which advocate Bayesian-like approaches to exposure assessment (Hewett et al, 2006).

Hawkins and Evans (1989) randomly selected 24 experienced OHs to estimate the distribution of exposures by leveraging experience and professional judgment on a single manufacturing process using toluene. Each OH estimated the mean, median, range, and 90th percentile exposures using only qualitative information about the process, tasks involving toluene, photos and information on the workforce. Next, the OHs were given limited historical exposure data and then estimated the mean, median, range, and 90th percentile exposures. The subjective judgments were then compared with an empirical distribution determined from 134 recent exposure measurements collected from the process. The authors concluded experts gave better estimates for the mean and 90th percentile than the median, and that their accuracy improved substantially when presented with the limited historical data. The study suggests that experienced industrial hygienists can provide good estimates of exposures.

Ramachandran and Vincent (1999) proposed a probabilistic Bayesian method for combining expert judgment, historical information about workplace conditions, and incomplete or limited measurements, so to determine a time and location exposure function. This Bayesian approach to exposure assessment evaluated the feasibility of reconstructing past exposures to inhalable nickel aerosol in a smelting operation (Ramachandran, 2001). The study used occupational experts to generate inputs for deterministic models

such as general ventilation without knowledge of exposure measurements. The authors used the output of the model to generate the exposure estimate and confidence. Many of the experts did have very similar professional experience which may have contributed to the correlation found in the study. However, the study did show that deterministic models may be a valuable tool for making exposure judgments. A Bayesian approach was utilized by Wild *et al*. (2002) to combine expert ratings and exposure measurements for epidemiological purposes with outcomes are similar to those obtained for similar reasons by Ramachandran (2001).

Walker *et al*. (Walker 2001, 2003) investigated the accuracy, certainty and correlation between expert judgments for non-occupational exposure to benzene. Qualitative information was given in a workshop setting to each expert to estimate the mean, inter-quartile range and 90% confidence interval. The studies indicated that as a group the expert's judgments were overall accurate and well correlated. The first study indicated that the experts were better with mean exposure than upper tail estimates. However after calibration, the judgments were more representative of truth for upper tail than mean exposure estimates. The mean exposure estimates appeared to be overconfident while the upper tail estimates were under-confident to a lesser degree. These studies also suggest that the collective judgments of the groups were correlated better with truth than the individual experts. The authors briefly referenced the heuristics and common biases proposed by Kahneman and Tversky but did not infer how they may be involved in the context of making exposure assessments.

Each of the studies mentioned showed that exposure judgments could be accurate in given applications. Several of the studies indirectly studied several possible determinants of exposure assessment accuracy. Ramachandran (2001) showed that by using physical exposure models, experts could improve exposure judgments. Several of the studies showed that accuracy of exposure judgments increase with increasing information and data presented (Hawkins and Evans, 1989; Walker *et al*., 2003). Studies used varying levels of "experts" which under careful inspection contains some of the elements fundamental to determinants affecting exposure assessment accuracy. Studies from psychology and other fields indicate that training and level of education can play a significant role in judgments accuracy illustrating the need to investigate training and educational interventions that can be used to enhance exposure judgment accuracy. Hewett (2006) and Ramachandran (2003) have highlighted a path for the use of Bayesian methods in the context of exposure assessment while studies in cognitive psychology highlight how humans are inherently Bayesian thinkers in some contexts (Gigerenzer, 1995; Zhu, 2006). These connections found throughout several fields of literature are positive indicators to utilizing a Bayesian construct with the AIHA exposure assessment strategy to better understand determinants affecting exposure judgment accuracy and designing methods to improve accuracy and removing bias.

PRELIMINARY STUDIES

Qualitative Exposure Judgments of Occupational Hygienists at a Large Manufacturing Facility

A pilot study was performed to evaluate agreement of exposure judgments between seven occupational hygienists (OHs) with different educational backgrounds, and lengths and types of professional experience. A focus was to begin identifying characteristics or determinants that could affect the accuracy and precision of exposure judgments, as well as to gain insight on data collection methods and determinants to be used in a larger and more comprehensive study.

The study was performed on seven different manufacturing tasks using different chemical agents at a large manufacturing facility. The task durations varied for each task and ranged between 30 minutes and 4 hours. All of the tasks involved some type of material transfer between containers and equipment. The chemical used in the tasks had a wide range of vapor pressures and occupational exposure limits (OELs). All of the OHs had previous experience conducting exposure assessments and controls for these types of tasks at this facility and other similar manufacturing facilities. Each participant was required to make exposure judgments using all of the available information and tools in accordance with routine exposure assessment documentation. The OH's were asked to predict for each task, the probability of the exposure distribution's 95th percentile falling into one of four categories defined in terms of percentages of the OEL found in the AIHA strategy. Task exposure judgments were made against the 8 hour OEL or Threshold Limit Value (TLV) published by the ACGIH (2003) regardless of the time duration of the task. Information utilized by the OHs included details describing the manufacturing processes, task description and duration, equipment, physical and chemical properties of each agent, engineering controls such as local exhaust ventilation, and Occupational Exposure Limits (OELs). However, they were not provided actual exposure sampling data for the selected tasks. The OHs documented their exposure judgments and determinants individually in a formatted spreadsheet. The rules of this exercise prohibited any discussions among the OHs regarding their individual exposure judgments to prevent influencing each other's judgment.

The OHs also had to observe some constraints, i.e., no category should have a 0% probability, and the sum of all categories must equal 100%. They were instructed to pick one category to have a higher percentage than all the other categories. If the OH could not pick the highest exposure category, then they could ask for additional process, task and chemical information until a judgment could be made. While the OHs were not permitted to discuss, comment or communicate regarding the details of the assessments or review sampling data for the types of tasks in the study, they were allowed to interview employees familiar with the task and equipment and view the operation.

Several potential determinants that might have influenced their exposure judgments were also documented: (a) years of exposure assessment experience (EA), (b) experience in mathematical exposure modeling

(Mod), (c) experience in exposure statistics (Stats), (d) experience in sampling specific agent, and (e) process experience. These variables are listed along with the various levels of each variable in Table I.II

Table I.II. Potential determinants of accuracy of exposure judgments by occupational hygienists used in pilot study.

Determinant variable	**Levels of variable**
Years of experience in exposure assessment	<1 years 1-3 years 3-5 years 5-10 years >10 years
Experience in mathematical modeling of exposures	1 = Never used a mathematical exposure model 2 = Used only simple exposure model (i.e. saturation vapor pressure) a few times 3 = Some experience using different exposure models 4 = Frequently use exposure models 5 = Have designed mathematical exposure models
Experience in exposure statistics	1 = Do not have a good understanding of normal and lognormal statistics 2 = Can do basic lognormal statistical calculations (GM, GSD) 3 = Familiar with calculating point estimate of arithmetic mean & 95%tile 4 = Familiar with calculating exceedance fractions and tolerance limits
Experience in sampling specific agent	1=never sampled it 2=observed sampling 3=collected <3 surveys on agent 4=collected 3 to 10 surveys on agent 5=greater than 10 surveys on agent
Process Experience	< 1 year of experience in this type of process / task 1-3 years of experience in this type of process / task 3-5 years of experience in this type of process / task 5-10 years of experience in this type of process / task >10 years of experience in this type of process / task

Table I.III. Regression models of the frequency of an OH being in the majority agreement category, f(MC) against years of exposure assessment experience (EA), experience in mathematical exposure modeling (Mod), experience in exposure statistics (Stats), and the sum of these three types of experience.

Statistical Model	*p*-value	R^2_{adj}

f(MC) ~ EA	0.030	0.57
f(MC) ~ Mod	0.393	0.00
f(MC) ~ Stats	0.088	0.37
f(MC) ~ EA + Mod + Stats	0.005	0.78

None of the self reported experience determinants appear to have a direct correlation with exposure judgment accuracy. However, when all self reported determinants of exposure assessment, modeling and statistics experience were added together, a linear correlation (R=0.78) with exposure judgment accuracy appeared to be statistically significant (p=0.005) Table I.III.

Data Interpretation of Limited Datasets by Occupational Hygienists (OHs)

Two pilot studies with two different groups of professional occupational hygienists were performed to explore each group's ability to interpret sampling data without the use of statistical tools. The objectives of the studies were to:

- Investigate the accuracy and bias associated with interpreting monitoring data without the use of statistical tools for several diverse groups of hygienists
- Qualitatively investigate the perspectives of hygienists utilizing a probabilistic framework to document exposure judgments using sampling data in the absence of workplace and basic characterization information.
- Qualitatively study whether training would increase accuracy and reduce bias of the group's judgments.

These studies were performed with groups of OHs at the American Industrial Hygiene Conference and Exposition (AIHCE06) and the Professional Conference of Industrial Hygiene (PCIH05) respectively held in Chicago, IL and Denver, CO. PCIH05 had 21 participants and AIHCE06 had 31 participants respectively. Participants had different education backgrounds and were employed in a variety of different business sectors including large and small company manufacturing, government and consulting. Each participant completed a Data Interpretation Test after a presentation was given to the group explaining the use and rules of the DIT. After each participant completed the first DIT, a short 30-minute training on lognormal statistics related to upper tail estimates and several rules of thumb to estimate the location of the 95th percentile were delivered to each group. A second DIT was administered to each participant following the 30-minute statistics and rule of thumb training. The "pre-training" DIT was used as the "post-training" and vice versa between PCIH05 and AIHCE06 to account for any DIT specific biases.

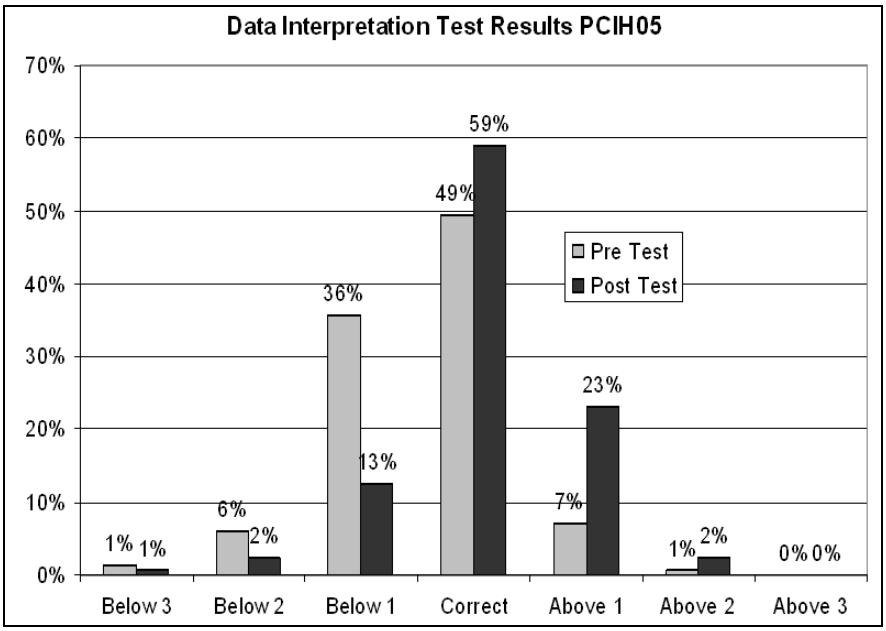

Figure 1.5 – Data Interpretation Test (DIT) results for a group of 20 occupational hygienist before (Pre Test) and after (Post Test) data interpretation training

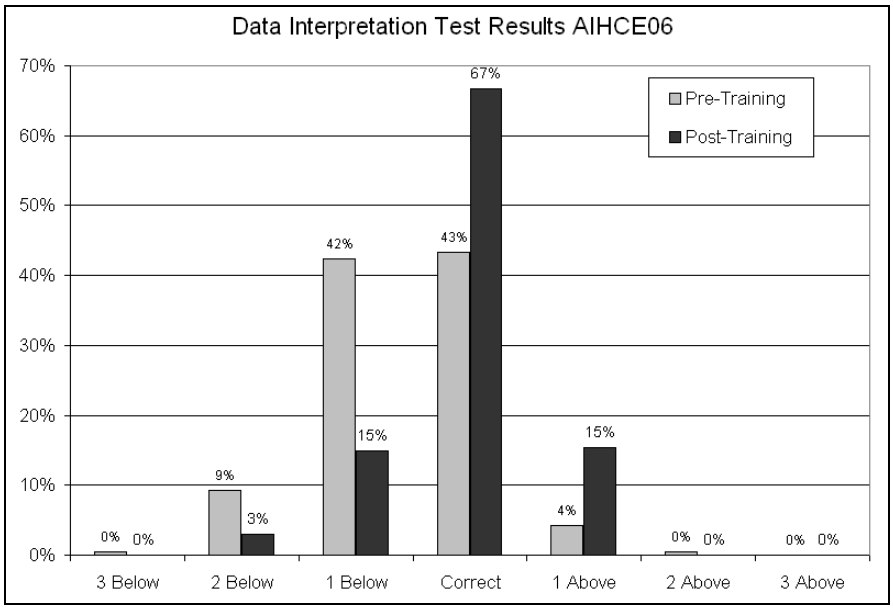

Figure 1.6 – Data Interpretation Test (DIT) results for a group of 31 occupational hygienist before (Pre Test) and after (Post Test) data interpretation training

The probability of selecting the correct category with DITs using the 4 AIHA categories purely by random chance is 25%. In all DITs it appears that participants were far more accurate than chance alone even without the rule-of-thumb training (49% and 43% in Figures 1.5 and 1.6 respectively). Both pre-training DITs indicated that judgments were biased low which means that hygienists were more likely to select a

lower exposure category than a higher exposure category when they were incorrect. We will refer to this bias as the "*lognormal is not normal*" bias in exposure assessment. After the statistics and rule-of-thumb training, there appeared to be a slight high bias in the first group and no significant bias in the second group. There was also a significant portion of judgments that were incorrect which indicates the presence of as yet to be identified heuristics and biases in exposure judgment making. The regression analysis studies planned may provide some insights into the judgment heuristics. This particular finding indicated a bias of underestimating exposure, which led to an additional research question. *Can statistics and data interpretation training reduce bias and increase accuracy in exposure judgments?* As a result of this new question inferred early in the preliminary studies, the research design was altered to include adding a training to see if accuracy and bias could be properly influenced toward truth.

RESEARCH DESIGN AND METHODS

Establishing a Decision Statistic

The following scenario and multiple-choice question was used to informally query several large groups of occupational hygienists so a commonly used decision statistic could be identified for this study. The scenario was delivered during each of the desktop data collections with participants coming from a wide range of employers including but not limited to government, consulting, private industry and academia.

"Assume that a worker performs a job 100 times over the next year and that a personal full shift sample will be collected each time the employee performs the job. What is the maximum number of samples out of 100 that can exceed the occupational exposure limit and still be considered acceptable without using a respirator or any additional controls?

A) 0 samples
B) 1 sample
C) 5 samples
D) 10 samples
E) 25 samples
F) 50 samples

Assume the chemical with acute and chronic health effects and is not a known human carcinogen. Please make a selection that describes the highest number exceeding the OEL and still considered acceptable."

The most common choice for the scenario by participants was 5 samples, which is analogous to utilizing the population 95th percentile as the decision statistic. This is consistent with several references that advocate the same upper tail interpretation of occupational exposure limits (Hewett 2001). Based on these informal surveys and documented references, this study utilizes the 95th percentile ($X_{0.95}$) as the "Decision Statistic" which defines a clear line of acceptable exposure to a majority of chemical agents. This approach to identify the decision statistic for current and prospective exposure assessments can be modified to

identify a decision statistic for retrospective exposure assessment and exposure matrices used for epidemiological studies of chronic disease agents. In such cases, a central tendency decision statistic would be more appropriate.

Creating Data Interpretation Tests

The Data Interpretation Test or DIT was developed to establish a metric to measure a participant's ability to estimate the probability of 95th percentile ($X_{0.95}$) being in each of the 4 AIHA categories for limited sampling datasets ($\leq$6 samples) without the use of statistical tools. A default OEL (Occupational Exposure Limit) of 100 ppm was selected for all DITs. Each data set was generated using a random sample generator or by using actual sampling data normalized to a 100 ppm OEL. Each DIT contained 8 different data sets which were distributed across the 4 AIHA exposure categories. Each DIT was administered with a short training which included the selection of the decision statistic. Participants were instructed to make computations manually or without the use of calculators or computers. The participants wrote the probability of the 95th percentile being in each of the 4 AIHA exposure categories. There was not strict time limit applied to the test however, all DITs were completed in less than 30 minutes. Multiple DITs with different data sets were created so each participant could be tested multiple times and reduce test specific biases. The number of samples in each dataset varied from 1 to 6 samples.

Data Interpretation Test (DIT) #1

OEL for all Data Sets 100								
	Sample Data Set #1	Sample Data Set #2	Sample Data Set #3	Sample Data Set #4	Sample Data Set #5	Sample Data Set #6	Sample Data Set #7	Sample Data Set #8
	21	5	5	68	1	10	7	3
	7		4		5	105	57	2
	53		23		2	20	34	22
	45		8		1	50		0
	16					10		4
	10					10		1

Make your judgments on the above Statistics Test Data in the following columns

	Data Set #1	Data Set #2	Data Set #3	Data Set #4	Data Set #5	Data Set #6	Data Set #7	Data Set #8
<10% OEL								
10-50% OEL								
50-100% OEL								
>100% OEL								

Figure 1.7. Example of a DIT (Data Interpretation Test) used during one of the desktop data collections

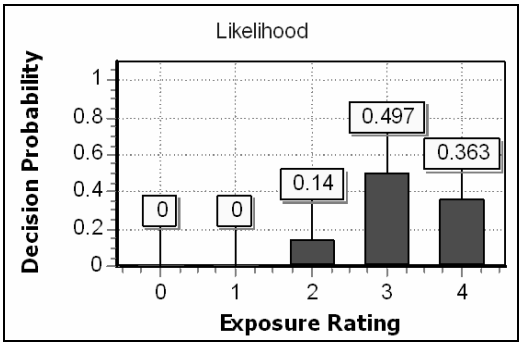

Figure 1.8. Likelihood probability chart for DIT Sample Data Set #1 using the IH Data Analyst software.

Each data set was analyzed using the IH Data Analyst software (Hewett et al, 2006) to identify which AIHA exposure category had the highest probability. Figure 1.8 illustrates the likelihood probability chart for Sample Data Set #1 from Figure 1.7. For each data set, a participant's highest category was compared to the results from the likelihood chart to determine whether the judgment was correct. The number of correct and the total number of data sets were used to calculate the DIT Metric using the following equation.

DIT Score = [(number of correct judgments) / (total number of judgments)] * 100

The DIT Score was calculated for each DIT taken by each participant and will be used as one of the determinants in the data analysis. The DIT Score does not account for the certainty of exposure assessment. To better account for certainty in judgments, the Score to test accuracy must also account for the probabilities in each category compared with the likelihood calculations. Since some of the participants only selected the most likely category and did not fill out probabilities for each data set, certainty cannot be adequately measured from the DITs.

Reference Judgments or "Truth" using Task Based Exposure Measurements

The probabilistic likelihood calculations using task-sampling data alone will be considered "truth". Each of the task exposure measurement data sets will be analyzed by calculating the probabilities of the $X_{0.95}$ falling into each of the 4 AIHA categories using Bayesian calculations discussed earlier (Hewett et al, 2006). This procedure will be applied to define "truth" in defining the most likely category and probabilities for each category. Each task will have 8 or more samples collected that will be used in "Truth" calculations. Information facilitating within and between worker variability calculations will also be collected for each task. Some samples collected may be below the limit of detection (LOD) will create censored data sets (Hewett et al, 2006). For tasks with censored datasets, the censored data analysis described by Hewett et al will be used to define "Truth". See detailed discussion of "Truth" in the "Data Analysis" Section

Desktop elicitation of Occupational Hygienist's Determinants and Subjective Exposure Judgments

Desktop elicitation exercises are designed to provide information that could be typically used to make exposure judgments. At the beginning of each session, a presentation was given to each group of participants reviewing the purpose of the study, procedures and rules to help reduce potential biases that could be introduced in the data collection. Each participant was asked not to talk, make gestures or noises that could bias other participants in any way. The data collections occurred on 5 different occasions with the number of participants ranging from 6 to 37. Data collections 1 and 2 did not include training on how to estimate $X_{0.95}$ falling into each of the 4 AIHA categories. In order to study the effect of training on judgments, a simple rule of thumb or Hewett Exposure Judgment Heuristic (HEJH) for the $X_{0.95}$ of lognormal data was introduced in data collections 3, 4 and 5.

Desktop data collections 1 and 2 used the following schedule:

- Overview presentation including procedures and rules for participation
- Data Interpretation Test (DIT)
- Video desktop judgment collection exercise for several tasks

Desktop data collections 3, 4 and 5 used the following schedule:

- Overview presentation including procedures and rules for participation
- First Data Interpretation Test (DIT)
- Video desktop judgment collection exercise for first set of tasks
- Training of participants on the Hewett Exposure Judgment Heuristic (HEJH) for the $X_{0.95}$ of lognormal data
- Second Data Interpretation Test (DIT)
- Video desktop judgment collection exercise for second set of tasks

Written descriptions of each task, chemical of interest, OEL to be used for the assessment, amount of material used in each task, chemical composition, chemical vapor pressure at normal conditions, equipment, ventilation controls and general work practices were given to each participant. A video of a worker simulating or actually performing each task was shown to each group of participants while the study leader narrated all of the written task information. A form containing all of the task information and tables to record sample data and probability exposure judgments was provided to each participant. Figure 1.9 illustrates an example of the data collection table provided.

OEL = 2 ppm	Initial Judgment	Sample #1	Sample #2	Sample #3	Sample #4	Sample #5	Sample #6	Sample #7	Sample #8
< 10%									
10-50%									
50-100%									
>100%									

Figure 1.9 - Example of a spreadsheet used by participants to record $X_{0.95}$ probability judgments and sample data given for a task.

Each participant was asked to estimate the probability of the $X_{0.95}$ being in each of the 4 AIHA categories given the information provided for the task. The participants were also asked to consider only the task duration described and not to extrapolate a full shift exposure by making assumptions about exposure for the rest of a shift. All participants were asked to adhere to the following rules when documenting exposure judgments.

- Record the sample result in the box with the Sample #
- Record the probability of $X_{0.95}$ being in each of the 4 AIHA exposure categories listed on the left most column
- One cell in each column must be higher than the other 3 cells indicating that the $X_{0.95}$ has the highest probability of being in that category
- The lowest probability for any one cell is 1% and therefore the highest probability for any one cell is 97%
- All 4 cells must add to 100%

The first judgment in the "Initial Judgment" column is considered a qualitative judgment since no sampling data was provided to the participants at that time. Once all of the participants had made their initial judgment after reviewing task information and the video, the first monitoring data point was verbally given to the group of participants. Once each participant had completed making the judgment for the 4 categories, the participants were instructed not to change any of the recorded probabilities. The participants were asked to write the value of the sample point on the form provided along with their probability judgments on the form provided. Then the second monitoring data point was revealed to the participants and the same procedure as before was repeated. This judgment elicitation proceeded until the last data point was given. The video of the task played for the duration of the elicitation exercise.

At the beginning of each data collection each participant was asked to record several personal attributes related to their education and self reported general experience with exposure assessment, modeling and statistics experience. A summary of an example determinant data collection sheet is found in Figures 1.10 and 1.11.

Location		Today's Date	
Print Name		Phone Number	
Current job title			
Years in current job type			
Undergraduate degree field	(Ex - BS Chemistry)		
Masters degree field			
Advanced degree field			
List certifications (CIH, CSP, IHIT, PE or other)			

	Please Circle Box that best applies and enter actual estimate				
Number of Years making exposure judgments	less than 1 year	1 to 3	3 to 5	5 to 10	greater than 10
Number of years since actively engaged in doing exposure assessments	less than 1 year	1 to 3	3 to 5	5 to 10	greater than 10
Estimate the total number of process-task WorkCHEC Exposure Assessments	less than 10	10 to 100	100 to 1000	1000 to 5000	> 5000
Career Air Sampling Experience	Collected less than 3 air sampling surveys	Collected 3 to 20 air sampling surveys	Collected 20 to 100 air sampling surveys	Collected 100 to 200 sampling surveys	Collected >200 air sampling surveys
Modeling Experience	Never used a mathematical exposure model	Used only simple exposure model (ie vapor pressure) a few times	Some experience using different exposure models	Frequently use exposure models	Have designed mathematical exposure models
Statistics Experience	Do not have a good understanding of normal and lognormal stats	Can do basic lognormal stats calcs (GM, GSD)	Familiar with calculating point estimate of 95%tiles or Exceedance Fractions	Familiar with calculating confindence intervals or tolerance limits for 95%tiles	

Figure 1.10. Participant's education and general experience determinant collection sheet.

For each task, participants recorded specific experience with each task and chemical. The purpose is to identify potential more detailed experience determinants.

Styrene Drum Charging Task	Please Circle Box that best applies for each row				
Years experience with exposure assessments on this type of task	Never	1 to 2	3 to 5	5 to 10	greater than 10
Approximate number of exposure assessments on this chemical	never assessed exposure to this chemical	1 to 2	3 to 5	5 to 10	greater than 10
Number of air sampling surveys performed on this chemical	never sampled it	oberved sampling	collected <3 surveys on agent	collected 3 to 10 surveys on agent	greater than 10 surveys on agent
Do you recall reviewing sampling data on this type of chemical task?	no	vaguely	yes	yes - on this exact task	

Figure 1.11. Participant's specific task and chemical experience determinant collection sheet.

Metrics to Compare Judgments with "Truth"

Properly developed metrics can help to better understand bias and more importantly identify heuristics that are used for exposure judgments and possible biases that result from their application. Consider that each judgment and "Truth" likelihood chart contains information on the most probable exposure category along with the judgment or monitoring data certainty. The judgment and certainty is based on the information

and data used to make the judgment or the exposure monitoring data used for the professional judgment and "Truth" likelihood chart. The "accuracy" of a given judgment can be expressed both in terms of the category with the highest probability (exposure control) being correct and the amount of probability overlap between the judgment and "Truth". A perfectly accurate judgment would be one where the judgment's highest category and probabilities in all other categories perfectly matched "Truth" (Figure 1.12a). A totally inaccurate and biased judgment is where the highest category does not agree with "Truth" and there is almost no probability overlap among all of the exposure rating categories (Figure 1.12b). The amount of agreement with "Truth" has two components; highest probability category and probability overlap of all categories.

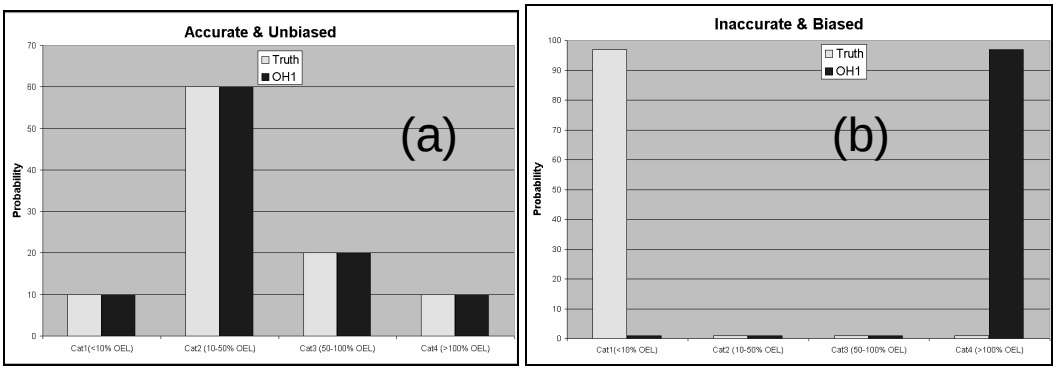

Figure 1.12. Charts of exposure judgments (OH1) versus "Truth" illustrating a perfectly accurate judgment (a) and a totally inaccurate and biased judgment (b).

This study proposes to analyze the accuracy of participant's judgments form interpreting information and data without statistical tools and the determinants or characteristics of participants including a rule of thumb training that may affect exposure judgment accuracy. This analysis will utilize several different metrics to compare judgments made with "Truth" and account for the relative accuracy against "Truth" on each judgment. "Truth" is defined in this study as the categorical probability likelihood distribution from sampling data collected on each of the tasks using validated sampling methods. If a metric does not properly account for varying levels of accuracy and certainty, the determinants that most effect accuracy may not be identified. Three different metrics have been used in other applications to compare probabilistic outputs.

Metrics for Comparing Exposure Judgments with "Truth"

There are several methods which could used to compare the exposure judgments provided by the participants with the sampling data obtained for each task or "Truth".

a. *Fractional Overlap (FO) -* For a given task judgment, one could compare the highest probability category provided by each participant with the highest likelihood category probability "Truth" calculated for the data set available for the judgment. We will refer to this

metric as the "Fractional Overlap" or FO using the calculation in Equation 1 for each judgment where $CatT_i$ and $CatJ_i$ are the probability in each ith exposure rating category for "Truth" and exposure judgment respectively.

$$\boxed{FO = \left(1 - \sum_{i=1}^{4} \left(CatT_i - CatJ_i\right)^2\right)} \text{ (Equation 1.1)}$$

b. *Percent Category Agreement (PCA) -* For a given task judgment, one could compare the highest probability category provided by each participant with the highest likelihood category probability "Truth" calculated for the data set available for the judgment using the formula in Equation 2. This calculation only uses the category number (1-4) with the highest probability in the "Truth" and exposure judgment charts and does not consider actual probabilities. HC_T and HC_J are the highest category identified by Truth and Judgment respectively.

$$\boxed{PCA = 1 - \left(\frac{\left|(HC_T - HC_J)\right.}{3}\right)} \text{ (Equation 1.2)}$$

c. *Probability Overlap (PO) -* To compare the exposure judgments of each participant with truth, one could compute a probability overlap defined as where $CatT_i$ and $CatJ_i$ are the probabilities in each ith category (1-4) for "Truth" and each judgment respectively. If the exposure judgments between a participant and "Truth" are identical, then the probability overlap is unity. The fractional overlaps for all combinations of participants and "Truth" across the range of task groups.

$$\boxed{PO = 1 - \sum_{i=1}^{4} \left|CatT_i - CatJ_i\right|} \text{ (Equation 1.3)}$$

Equations 1-3 illustrate different approaches to calculate the "accuracy" of a given judgment using only the highest probability category or the probability overlap across all four exposure categories. Exposure judgments can be accurate or biased by overconfidence, underconfidence, or from problems inherent to the judgment heuristics used. Figure 1.13 (a-e) illustrates several different biases that can occur in exposure judgments compared with "Truth". Consider two judgments made on the same task by different hygienists may both have selected the correct exposure category compared with truth but may have very different probability overlap with "Truth" (see Figures 1.13a and 1.13b). If a judgment is overconfident, it may contain a higher probability in a given exposure category than what is warranted given all of the data and information. In order to thoroughly study the accuracy of exposure judgments it is important to consider how accuracy of judgments aligns with the conventional thinking of occupational hygiene professionals. If

two judgments are one category above and one category below “Truth” respectively, the judgment that is below “Truth” may be determined to be less accurate because it is potentially underprotective (Figure 1.13c and 1.13d).

Exposure assessment professional need a quantitative method or tool which can be used to provide feedback when exposure judgments are made in a probabilistic context. Equations 1.1, 1.2 and 1.3 can be used to evaluate the relative accuracy of judgment when sampling data is available. No metric is perfect since sampling data may be limited compared to a professional judgment. Where there is a robust professional judgment for a given job or task, there would need to be adequate sampling data to utilize the metrics. In either case, the use of exposure accuracy metrics could provide an important feedback mechanism to better calibrate exposure assessors ultimately leading to better protection of workers.

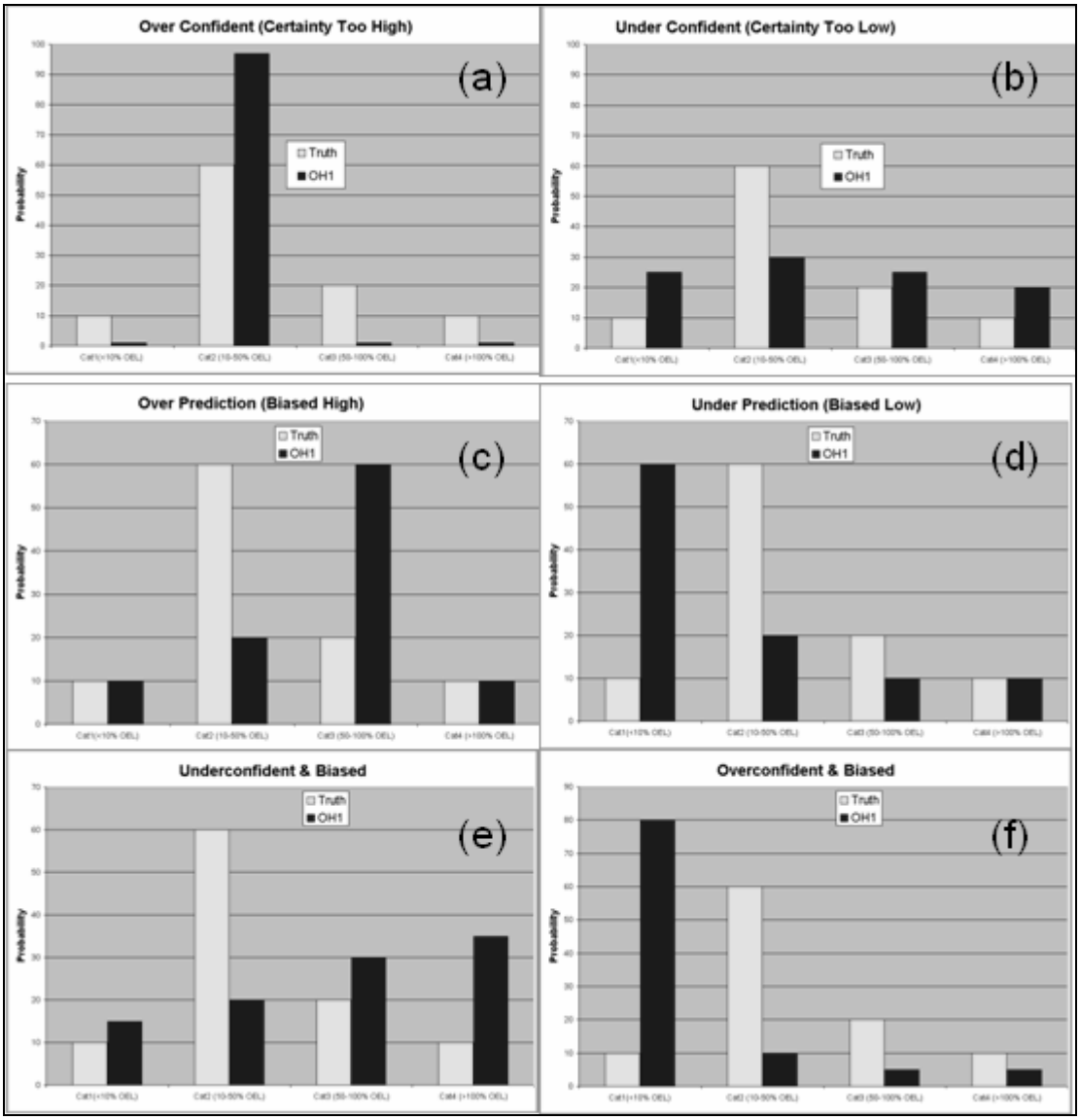

Figure 1.13. Charts of exposure judgments (OH1) versus "Truth" illustrating various possible judgment biases including (a) Overconfidence, (b) Underconfidence, (c) Over prediction / Biased High, (d) Under prediction / Biased Low, (e) Underconfidence with Bias, (f) Overconfident with Bias.

CHAPTER 2 - Occupational Exposure Decisions: Can Limited Data Interpretation Training Help Improve Accuracy? *Annals of Occupational Hygiene*, 1-14 (2009)

Accurate exposure assessments are critical for ensuring that potentially hazardous exposures are properly identified and controlled. The availability and accuracy of exposure assessments can determine whether resources are appropriately allocated to engineering and administrative controls, medical surveillance, personal protective equipment and other programs designed to protect workers. A desktop study was performed using videos, task information and sampling data to evaluate the accuracy and potential bias of participants' exposure judgments. Desktop exposure judgments were obtained from occupational hygienists for material handling jobs with small air sampling data sets (0-8 samples) and without the aid of computers. In addition, data interpretation tests were administered to participants where they were asked to estimate the 95th percentile of an underlying lognormal exposure distribution from small data sets. Participants were presented with an exposure data interpretation or rule-of-thumb training which included a simple set of rules for estimating 95th percentiles for small data sets from a lognormal population. Data interpretation test were given to each participant before and after the rule-of-thumb training. Results of each data interpretation test and qualitative and quantitative exposure judgments were compared with a reference judgment obtained through a Bayesian probabilistic analysis of the sampling data to investigate overall judgment accuracy and bias. There were a total of 4,386 participant-task-chemical judgments for all data collections: 552 qualitative judgments made without sampling data and 3,834 quantitative judgments with sampling data. The data interpretation tests and quantitative judgments were significantly better than random chance and much improved by the rule of thumb training. In addition, the rule of thumb training reduced the amount of bias in the data interpretation tests and quantitative judgments. The mean data interpretation test % correct scores increased from 47% to 64% after the rule-of-thumb training ($p<0.001$). The accuracy for quantitative desktop judgments increased from 43% to 63% correct after the rule-of-thumb training ($p<0.001$). The rule of thumb training did not significantly impact accuracy for qualitative desktop judgments. The finding that even some simple statistical rules of thumb improve judgment accuracy significantly suggests that hygienists need to routinely use statistical tools while making exposure judgments using monitoring data.

INTRODUCTION

Exposure assessment is at the core of environmental and occupational hygiene practice, and thus, a critical skill of occupational hygienists is making accurate exposure judgments to ensure that workers are properly protected and resources are efficiently utilized. This study aims to investigate (a) the accuracy and bias of exposure judgments using the American Industrial Hygiene Association (AIHA) exposure assessment

strategy with limited monitoring data for a specific set of exposure tasks, and (b) the impact of simple data interpretation training on the accuracy and bias in exposure judgments.

Exposure judgments are typically made across a continuum of available information and exposure data (Rock, 1986, Kromhout et al, 1987, Hawkins and Evans, 1989, Teschke et al, 1989, Macaluso, 1993, Cock et al, 1996, Freisen et al, 2003, Ramachandran et al, 2003, Ramachandran, 2008). Many exposure judgments are made with very limited monitoring data, such as retrospective exposure judgments for epidemiology studies and prospective exposure judgments for planning future manufacturing operations. In this paper, exposure judgments made without direct exposure monitoring data are referred to as *qualitative* and judgments made with monitoring data are *quantitative*. When limited or no sampling data is available, occupational hygienists typically review basic characterization information, leverage surrogate data, and use what can be referred to as "professional judgment" to arrive at a qualitative exposure judgment (Ignacio and Bullock, 2006). As expected, such professional judgments are often subjective, resulting in exposure judgments having a wide range of accuracy, depending on many factors (Kromhout et al, 1987, Hawkins and Evans, 1989, Teschke et al, 1989, Macaluso, 1993, Cock et al, 1996, Freisen et al, 2003, Walker et al, 2001, Walker et al, 2003, Ramachandran et al, 2003).

Several well known judgment strategies use a method of initially classifying workers into similar exposure groups (SEGs) based on observation of a task or group of tasks in a process (Corn and Esmen, 1979, Mulhausen and Damiano, 1998, Ignacio and Bullock, 2006). An SEG is a group of workers having the same general exposure profile for the agents(s) being studied because of the similarity and frequency of tasks they perform, the materials and processes with which they work, the administrative and engineering controls employed and the similarity of the way they perform the tasks. Occupational hygienists review the workforce, materials, exposure agents, tasks, work practices, equipment, existing administrative and engineering exposure controls, and then identify exposure groups that will be assessed, and possibly controlled, depending on the exposure judgments or exposure control category selected. The exposure judgment for any prospective SEG requires the selection of an occupational exposure limit (OEL) and a judgment by the hygienist about where the SEG exposure decision statistic falls in relation to the OEL. Chapter 5 in "A Strategy for Assessing and Managing Occupational Exposures", Third Edition contains a good discussion on OEL selection. The selection of a decision statistic is an important element for how judgments are performed and exposure controls are managed (Ramachandran, 2008). Using exposure assessment and control categories defined in the AIHA Exposure Assessment and Management Strategy, qualitative judgments can be documented in fractions or multiples of the selected health based or compliance based OEL (Mulhausen and Damiano, 1998, Ignacio and Bullock, 2006). The AIHA exposure control categories used in this study are illustrated in Table I.I. Judgments are made by identifying the exposure control category in which the 95th percentile of the exposure distribution is most likely located for

a given job or task. A judgment can be documented for each SEG, which can represent a single task that may be short in duration or may represent a group of tasks that comprise a full-shift exposure.

Table II.I: AIHA exposure category rating scheme. A Similar Exposure Group (SEG) is assigned an exposure rating by comparing the 95th percentile exposure ($X_{0.95}$) of the exposure distribution with the full-shift time-weighted average (TWA) Occupational Exposure Limit (OEL) or Short-Term Exposure Limit (STEL) to determine in which category it most likely falls (Mulhausen, 1998, Ignacio, 2006).

AIHA Exposure Rating	Proposed Control Zone Description	General Description	AIHA Recommended Statistical Interpretation
1	Highly-controlled (HC)	95th percentile of exposures rarely exceeds 10% of the limit.	$X_{0.95} \leq 0.10$ OEL
2	Well-controlled (WC)	95th percentile of exposures rarely exceeds 50% of the limit.	$0.10 \text{ OEL} < X_{0.95} \leq 0.5$ OEL
3	Controlled (C)	95th percentile of exposures rarely exceeds the limit.	$0.5 \text{ OEL} < X_{0.95} \leq \text{OEL}$
4	Poorly controlled (PC)	95th percentile of exposures exceeds the limit.	$\text{OEL} < X_{0.95}$

Expert Judgment

Professional judgment studies across many fields have identified similarities across professionals that are likely to be present in hygienists (Connolly, 2000, Gilovich et al, 2002). Studies investigating the accuracy of exposure judgments have shown that exposure assessment professionals tend to be more accurate than untrained non-professionals. However, some studies show that exposure judgments made by professionals may not always be highly accurate but can become more accurate with more data and training (Kromhout et al, 1987, Hawkins and Evans, 1989, Teschke et al, 1989, Macaluso, 1993, Cock et al, 1996, Freisen et al, 2003, Walker et al, 2001, Walker et al, 2003, Ramachandran et al, 2003). These studies focused on some measure of agreement between assessors or with monitoring data, but few studies looked at systematic biases in professional judgments and potential causes of the bias. Walker et al (2001, 2003) reported that environmental exposure assessments made by experts appear to have systematic bias and suggested that heuristics may be playing a role, as found in other professions.

Judgments made in medicine, psychology, law and other fields utilize available information and data using similar high-level cognitive processes as those used to make an exposure judgment. Extensive work in the field of psychology indicates that heuristics or simple decision rules make the decision process efficient but can lead to specific biases according to the heuristics or decision rules used (Kahneman et al, 1982, Griffin

and Tversky, 1992, Gilovich et al, 2002). A quote from Nobel Laureates Kahneman and Tversky illustrates this finding: *"In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction. Instead, they rely on a limited number of heuristics which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors"* (Kahneman et al, 1982). In their experience teaching and in applied studies, they found that both students and trained professionals tended to make common mistakes when interpreting data. Kahneman and Tversky proposed three heuristics that help explain the underlying mechanisms. These, now famous, heuristics are (1) availability, (2) anchoring and adjustment, and (3) representativeness.

After a review of professional judgment studies in other fields, one may conclude that hygienists also use common heuristics when making exposure judgments (Walker et al, 2001, Kahneman et al, 1982, Gilovich et al, 2002, Otway and von Winterfeld, 1992, Connolly, 2000, Griffin and Tversky, 1992, Gigerenzer, 1999, Baron, 2000). Since hygienists periodically need to make quick judgments with limited information and data, they rely on simple processes and heuristics to efficiently make judgments. One could imagine several instances where common heuristics would be utilized when making exposure judgments.

1. Judgments made using only the *available* information in memory is an example of using the *availability* heuristic. If an exposure judgment needs to be made quickly, a hygienist may use only the information that comes to mind rather than spending additional time and resources collecting more information and data. In this case, the judgment would only be as good as the information available in memory and, therefore, could be prone to memory bias.
2. Judgments can also made using the *anchoring and adjustment* heuristic by focusing only on one piece of information or data point without taking all information and data into consideration. Again, the quality of the judgment will be dependent on the information or data used to *anchor* the judgment.
3. An example of the *representativeness* heuristic could be seen where a hygienist fails to understand the most likely *representative* distribution for a population, such as using a normal distribution when the population is most likely lognormal. In this case, the hygienist would likely underestimate an upper tail decision statistic, resulting in an underestimate of exposure. One can imagine many other scenarios where a hygienist could use a combination of these and other heuristics to make exposure judgments.

A few authors have proposed methods for combining qualitative information with quantitative data using probabilistic methods to arrive at an integrated judgment in an exposure category (Wild et al, 2002, Hewett et al, 2006). A powerful attribute of this type of approach is exposure judgments can be defined in terms of the probability of the decision statistic falling in each exposure category. Hewett et al. illustrated the use of Bayesian statistical methods in the context of the AIHA strategy to combine qualitative and quantitative exposure judgments with monitoring data for a given SEG. This approach provides a transparent method

for incorporating the relative certainty of the information or data used to produce a judgment probability chart (Figure 2.1).

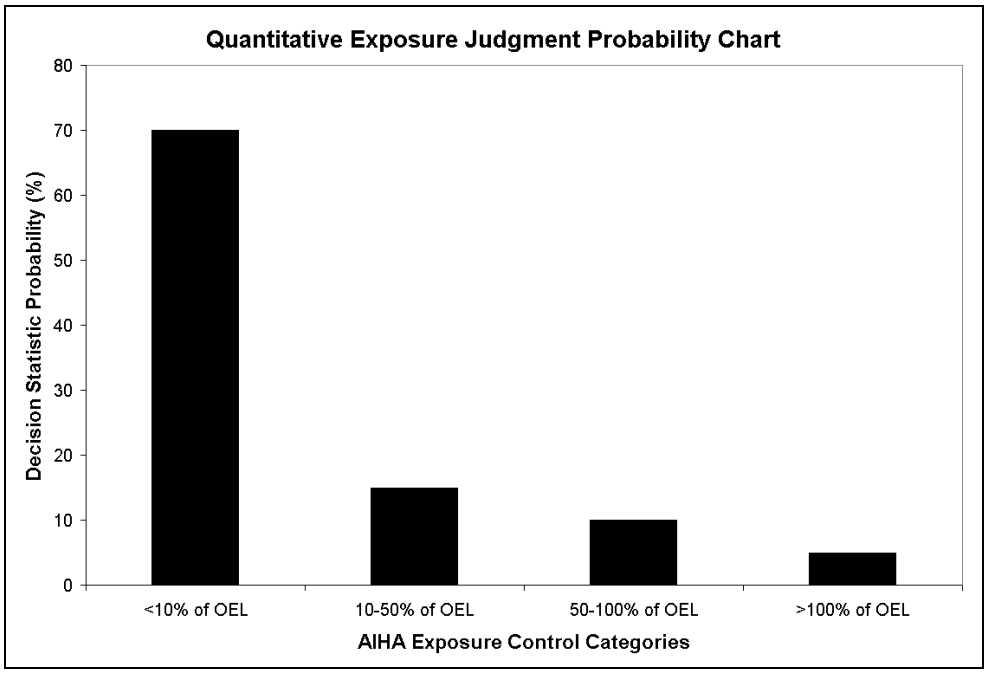

Figure 2.1: Example qualitative exposure judgment probability chart illustrating an occupational hygienist's exposure judgment given the information and data available. This chart shows that the hygienist is highly confident the 95th percentile falls into Category 1 – <10% of the OEL

The Bayesian approach applied to exposure judgments presents a structured method for utilizing qualitative judgments (Prior) and monitoring data (Likelihood) to create an integrated judgment (Posterior). The AIHA model integrated with Bayesian methods also provides a very powerful construct to test exposure judgment accuracy which may be based on qualitative exposure judgments or outputs from an exposure model. These exposure judgments for various scenarios can be directly compared with the Bayesian likelihood chart based on monitoring data to test the accuracy and bias for a group of assessments (Figure 2.2). By systematically collecting various attributes of the assessments or exposure models, this method can help identify factors that impact the accuracy of exposure judgments and provide insight for specific training or follow up (Hewett et al, 2006).

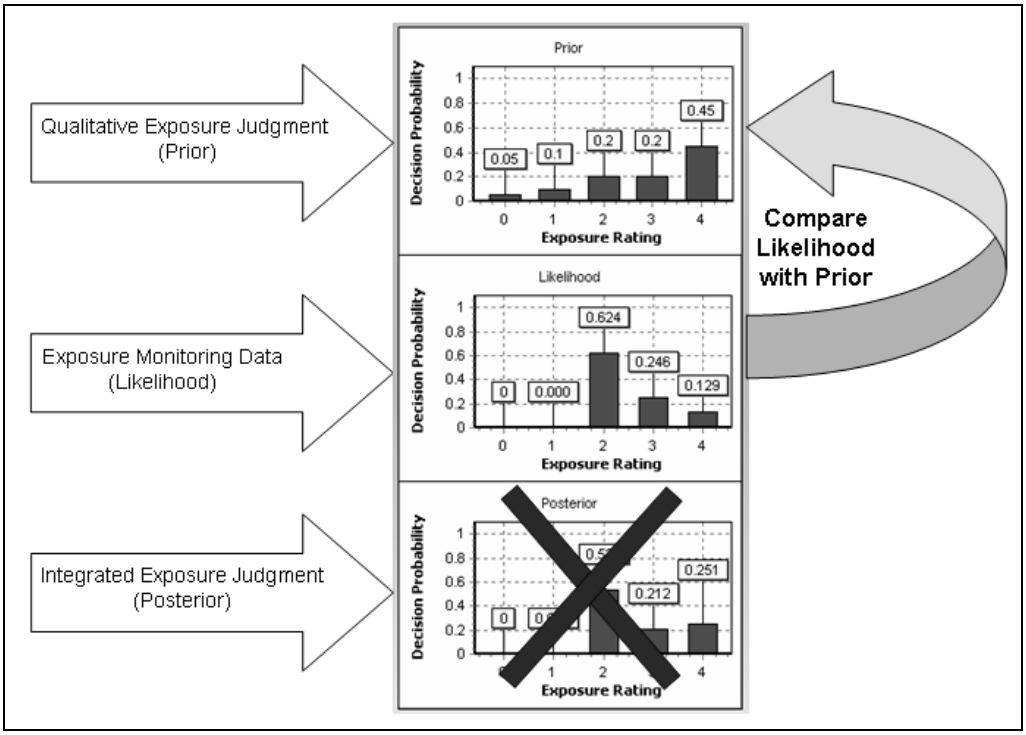

Figure 2.2: Bayesian integrated AIHA strategy used to test exposure judgment accuracy. The figure Illustrates a method for utilizing the Bayesian integrated AIHA Strategy to compare exposure monitoring data analysis (Likelihood) with exposure judgments (Prior) made by an occupational hygienist for a given SEG.

METHODS

Desktop Study Overview

A desktop study using videos and written and oral information was designed to study the accuracy and potential biases of exposure judgments for several tasks commonly found in industry. Participants made consecutive exposure judgments with increasing monitoring data while seated at desktops in conference rooms without the use of a computer or other statistical tools. Those judgments were then compared to a reference decision calculated for the task data using Bayesian statistical analysis (Hewett et. al. 2006) in order to characterize their accuracy. During each data collection, exposure judgments were collected for a subset of tasks before and after data interpretation training in order to evaluate the impact of that training on judgment accuracy. Participants were also given pre- and post-training standardized data interpretation tests (DIT) in order to further characterize each individual's ability to interpret a small data set (1 to 6 measurements) without the use of a computer or other statistical tools.

Subjects were asked to participate at several annual industrial hygiene conferences. Participation was solicited from conference attendees with various levels of experience making exposure assessments using AIHA exposure assessment strategy categories. Participants were given a brief description of the desktop study and instructed that none of the participants in the study or their current employer would be personally identified in any publication related to the study. All participants were asked not to disclose any of the information used in the study to avoid influencing future participants. No compensation was provided, and participation was voluntary.

Participants gathered in a conference room where the desktop study took place. The judgment elicitation sessions lasted between two to four hours depending on the number of exposure jobs reviewed. The study personnel presented an overview of the research goals, after which participants were instructed on the decision statistic to be used and procedures for providing probabilistic exposure judgments using the AIHA exposure categories in the data forms given to them. Information was collected on the participants' educational background and their professional expertise in various aspects of occupational exposure assessment (Table 2). The participants were then administered a pre-training DIT that measures the ability of the participant to estimate, based on a small monitoring data set, the probability of the 95th percentile of the exposure distribution being located in each of the four AIHA exposure control categories.

Table II.II: Summary of certifications, education and experience determinants for all participants (N=75)

	Yes	No			
CIH	61%	39%			
	Yes	No			
CSP	28%	72%			
	High School or Associates Degree	Bachelors Degree in EHS or Science	Masters Degree in EHS or Science	PhD in EHS or Science	
Highest Degree Achieved	15%	24%	59%	3%	
	less than 1 year	1 to 3	3 to 5	5 to 10	greater than 10
Number of Years making exposure judgments	17%	9%	9%	20%	44%
	less than 1 year	1 to 3	3 to 5	5 to 10	greater than 10
Number of years since actively engaged in doing exposure assessments	60%	19%	7%	8%	7%
	less than 10	10 to 100	100 to 1000	1000 to 5000	> 5000
Estimate the total number of job-tasks documented using AIHA strategy	31%	24%	35%	7%	4%
	Collected less than 3 air sampling surveys	Collected 3 to 20 air sampling surveys	Collected 20 to 100 air sampling surveys	Collected 100 to 200 sampling surveys	Collected >200 air sampling surveys
Career Air Sampling Experience	9%	12%	16%	24%	39%
	Never used a mathematical exposure model	Used only simple exposure model (ie vapor pressure) a few times	Some experience using different exposure models	Frequently use exposure models	Have designed mathematical exposure models
Modeling Experience	25%	39%	33%	1%	1%
	Do not have a good understanding of normal and lognormal stats	Can do basic lognormal stats calcs (GM, GSD)	Familiar with calculating point estimate of 95%tiles or Exceedance Fractions	Familiar with calculating confindence intervals or tolerance limits for 95%tiles	
Statistics Experience	13%	55%	9%	23%	

Following the pre-training DIT, participants were shown videos of a worker performing a task that had potential exposure to the chemical of interest. The eight tasks selected for this study included various types of material handling of liquids from drums or containers and cleaning tasks. The first judgment was made by each participant without any sampling data given, representing the initial or *qualitative* judgment. Subsequent *quantitative* judgments were made by participants as they were given a single personal sample data point, one at a time. Participants were not allowed to use computers or calculators but were able to make hand calculations on the data collection sheets provided.

After this, the participants were given a targeted data interpretation training that provided rules to estimate the mostly likely AIHA exposure control category. Simple "rules of thumb" or heuristics were developed that could can be applied easily to small data sets for estimating in which control category the 95th percentile most likely falls. The rules of thumb are presented in Appendix I and require, at most, four calculations that can be easily performed on paper or in one's head. Following this training, we again administered a different DIT and then asked the participants to provide exposure judgments on a different set of video desktop tasks. A flowchart describing the sequence of procedures followed during each judgment elicitation session is illustrated in Figure 2.3.

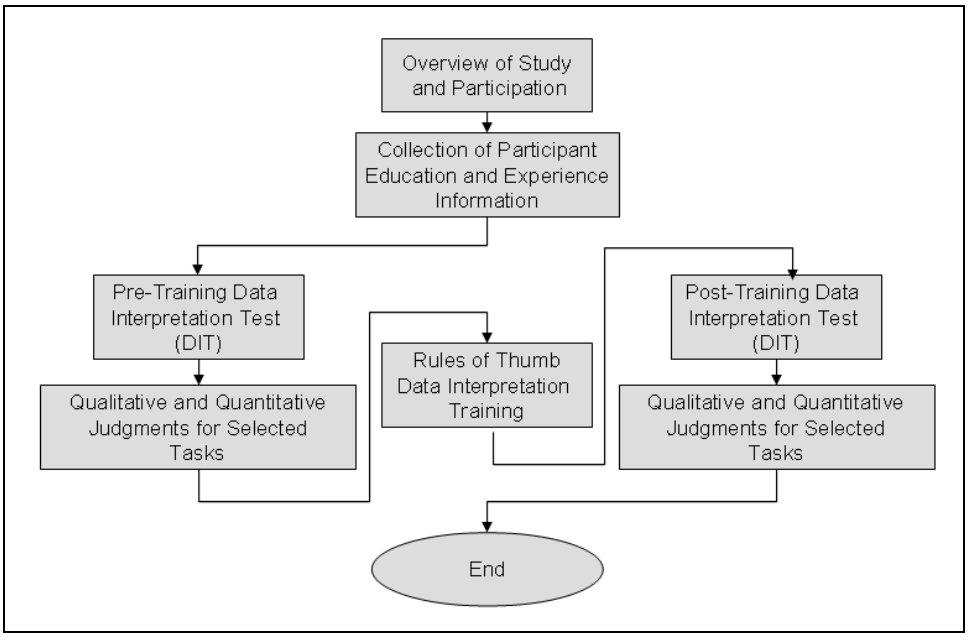

Figure 2.3: Flow chart of procedures for judgment desktop elicitation sessions

Calculating the "Reference" Judgment

To assess the accuracy of the hygienists' judgments made by hygienists, they need to be compared against a "reference" or "correct" judgment. The reference judgment is defined as one that is predicted by estimating the probability of the true 95th percentile falling into each of the four AIHA exposure categories given all of the available monitoring data. This is known as the likelihood distribution, which represents the relative probability of observing this set of data given all possible specific combinations of geometric mean (GM) and geometric standard deviation (GSD) (Hewett et al, 2006).

For each of the eight tasks, the number of samples used in quantitative judgments was less than the full data set used to calculate the reference judgment (Table II.III). For task data sets that contained one or more values less than limit of detection (LOD), censored data analysis using maximum likelihood estimation (MLE) method was used to select the AIHA categories that had the highest probability for the 95th percentile (Hewett et al, 2006, Hewett and Ganser, 2007).

Table II.III: Summary of sampling data used for reference calculations

Task	Samples Used for Quantitative Judgments by Hygienists	Total Number of Samples Used in Calculation of Reference Judgment	Number of Employees Sampled
1*	1-8	13	9
2	1-5	9	6
3	1-5	8	6
4*	1-5	14	10
5	1-7	8	5
6*	1-8	10	8
7	1-8	13	8
8*	1-6	9	7

** The censored data analysis using maximum likelihood estimation (MLE) method was used to determine the reference judgment because one or more data points were less than detection limit (Hewett et al, 2006, Hewett and Ganser, 2007)*

Data Interpretation Tests - DITs

A group of DITs were designed to better understand the accuracy and bias when estimating the decision statistic for a small data set without the use of computer-based statistical tools or any specific information about a job, task or chemical. It is possible that bias in estimating exposure with exposure data may also be present with qualitative exposure judgments where sampling data is not present. Each DIT had a total of eight data sets with the number of samples ranging from one to six samples. Data sets for each DIT were either created from actual sampling data normalized to an OEL of 100 ppm or by using a spreadsheet that randomly selected data from a lognormal distribution with a geometric standard deviation of 2.5 and various geometric means. Each participant was instructed to review each data set and estimate the probability of the 95th percentile falling in each of the four exposure categories, based on an OEL of 100 ppm. The participants were asked to ensure that the probabilities for each data set total 100% and that one category have the highest probability. Figure 2.4 illustrates a DIT that has eight data sets with an example judgment for Data Set #7 which has three samples. Reference judgments were calculated for each dataset for comparison to the exposure judgments made by participants.

Statistics Test Data

Enter your name:

OEL for all Data Sets
100

Sample Data Set #1	Sample Data Set #2	Sample Data Set #3	Sample Data Set #4	Sample Data Set #5	Sample Data Set #6	Sample Data Set #7	Sample Data Set #8
7	55	15	4	2	23	112	1
17		12		1	41	86	2
20		23		8	12	72	10
24		9		1	8		1
37					18		4
40					36		

Make your judgments on the above Statistics Test Data in the following columns

	Data Set #1	Data Set #2	Data Set #3	Data Set #4	Data Set #5	Data Set #6	Data Set #7	Data Set #8
1-10% OEL							1	
10-50% OEL							1	
50-100% OEL							18	
>100% OEL							80	
Check	0	0	0	0	0	0	100	0

Have you ever taken this statistical test before?	Yes	No
If yes, how many times?		

Figure 2.4: Example Data Interpretation Test with eight data sets

The category with the highest probability was compared to the reference judgment highest category for each data set. A DIT score was then calculated for each participant's DIT using the following formula in Equation 2.1:

$$DIT\ Score = \frac{Number\ of\ Correct\ Judgments}{Number\ of\ Data\ Sets\ Evaluated\ (\mathbf{8})} \tag{2.1}$$

Pre- and Post-Training Task Judgment Data Collections

Videos of a worker performing all or part of an actual exposure task were shown to each study group. In addition, for each of the tasks, written basic characterization information was given, including amounts of materials and mixture concentrations, physical properties of exposure chemical, approximate duration of each task, equipment and ventilation. Participants were allowed to question the proctor directly, but were asked not to communicate with other participants to prevent potentially influence. The participants were asked to document each judgment as a probability of the 95th percentile (decision statistic) falling in each of the four exposure categories. This probability judgment method was the same method used for documenting judgments in the DITs. Participants were first asked to document their "Initial Judgment", which is considered qualitative because no exposure data was yet available to participants. When all participants completed the qualitative judgment, the participants were given the first sample point and asked to document their next judgment, which would be considered quantitative. All participants were instructed not to go back and change any prior judgments throughout the data collection process. This

process continued until all available sample data for each task was utilized in the judgments. The same judgment elicitation process was used before and after training. There were 8 task and chemical combinations used in the study, all of which used different chemicals. The judgments made about each task were only for the task duration and not for the full-shift. If a task lasted 15 minutes, participants were asked to estimate the exposure during the 15-minute period.

Each judgment was compared to the "reference" standard for the task, and a category judgment score (CAT Score) was calculated based on the percentage of correct judgments made by a participant for a given task.

$$CAT\ Score = \frac{Number\ of\ Correct\ Judgments}{Total\ Number\ of\ Judgments} \tag{2.2}$$

RESULTS AND DISCUSSION

Data Interpretation Tests and Training

The Data Interpretation Test (DIT) was designed to help study how hygienists interpret small data sets ranging from one to six samples. The output of the judgment is constructed in probabilities of the 95th percentile falling into each of the four AIHA categories. This provides a transparent method for incorporating the level of certainty for a given judgment. The DIT does not specify a chemical and assumes that all data is of proper duration for comparison with the given OEL of 100 ppm. Therefore, a DIT score provides a measure of a hygienist's ability to estimate the 95th percentile of lognormal data. Since each judgment is made of four categories, the DIT score (percent of correct judgments) expected from random chance is 0.25 or 25%. As a whole, study participants did better than random chance, with mean pre- and post-training DIT scores for the group of 0.47 and 0.64 respectively (Table 4).

Table II.IV: Summary of t-tests comparing accuracy of pre- and post-training DIT scores to random chance test score of 0.25

	N	Mean	Standard Deviation	95% Confidence Interval	T	P value
Pre Training	41	0.47	0.23	(0.40, 0.55)	6.07	<0.001
Post Training	41	0.64	0.17	(0.59, 0.70)	14.9	<0.001

The percent of participant DIT scores above 50% correct increased from 44% before training to 90% after training (Figure 2.5). The 95% lower bounds for before and after training were 0.40 and 0.59 respectively,

which are both well above what would be expected from random chance (0.25). The upper bound of the mean for before training of 0.55 was below the lower bound of 0.59 for the mean after training, indicating that the rule-of-thumb training provided a statistically significant positive impact to data interpretation accuracy (Table II.IV). Before the rule-of-thumb training, about 24% of DIT scores were at or above 75% correct; after training, the percentage of DIT scores at or above 75% correct jumped to 46% (Figure 2.5).

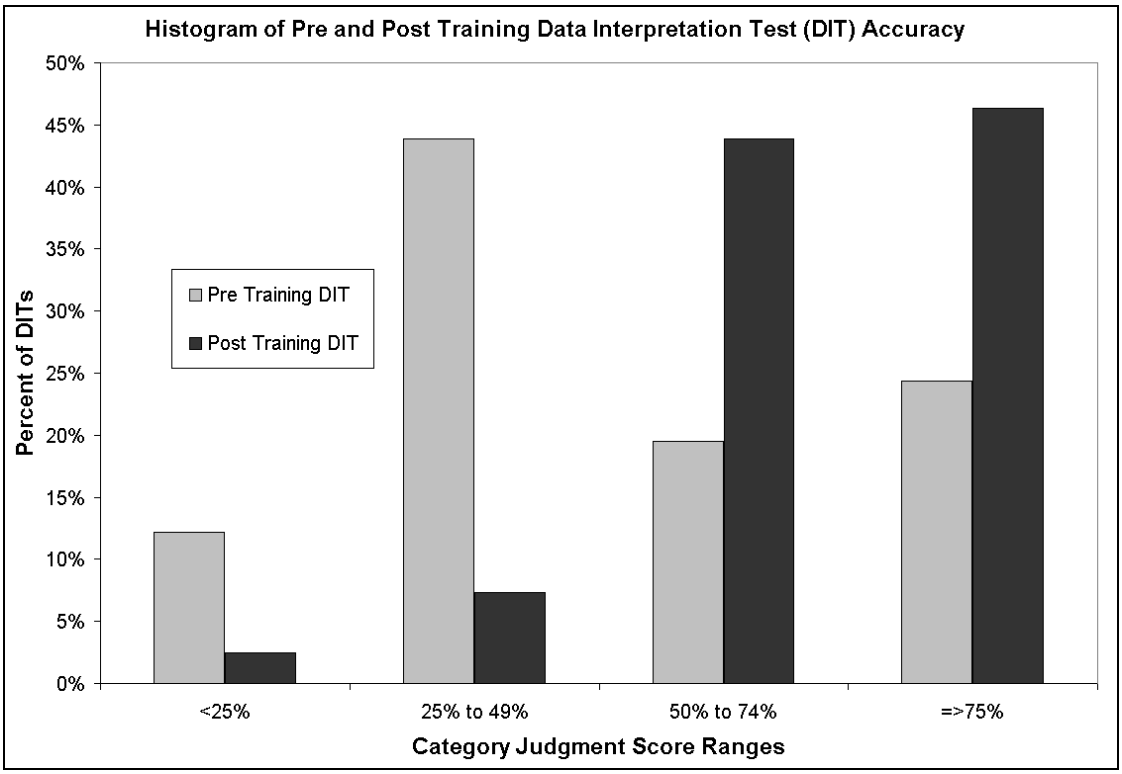

Figure 2.5: Histogram of all DIT scores for pre and post training

The DIT score is one measure of accuracy for data interpretation judgments but does not indicate whether or not bias is present across the judgments. The individual DIT judgments were compared to the reference and plotted in a histogram to help illustrate potential bias occurring across all of the judgments pre and post training (Figure 2.6). The DIT judgments made pre training appear to be biased low: ~38% of DIT judgments were below the Reference category, while only ~15% of DIT judgments were above it. This finding could be due to anchoring bias or base-rate error from improperly utilizing the Representativeness heuristic for normal rather than lognormal data (Otway and von Winterfeld, 1992, Kahneman and Tversky, 1979, Griffin and Tversky, 1992). The rule-of-thumb training appeared to shift the bias to slightly overestimate exposure for small data sets: ~10% of DIT judgments were below the Reference category while ~25% of DIT judgments were above the Reference category. There were a total of four different DITs with different data sets rotated between pre and post training to minimize any specific test effects.

Given the shift of bias from under-estimation to over-estimation, it appears that the DIT training had a significant and positive impact on DIT judgments.

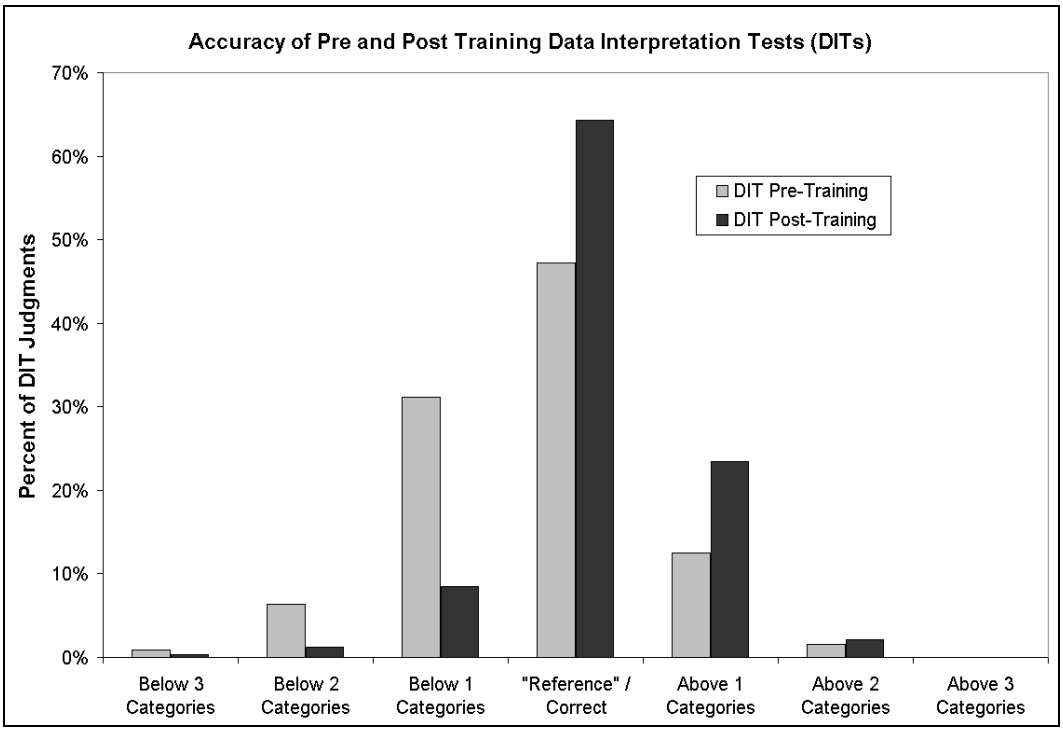

Figure 2.6: Percentage of all pre and post training DIT judgments above, below and "Reference" categories (N=82)

Qualitative Judgments and Training

The qualitative judgments were made after the basic characterization information and videos were reviewed but before any sample data were made available to participants. As with the DIT scores, the probability of a correct judgment due to random chance is 0.25 or 25%. Table 2.5 shows that the mean of all pre- and post-training qualitative scores for the group were 0.32 and 0.29, respectively. The 95% lower bound for the pre training was 0.27, slightly above what is expected from random chance. The 95% lower bound for after training was 0.23, which is slightly below what is expected from random chance. It appears that the pre-training qualitative judgments were above random chance while the post-training score may not be different from random chance. This indicates that the rule-of-thumb training did not positively impact the qualitative judgment accuracy.

Table II.V: Summary of t-Tests comparing accuracy of pre-and post-training qualitative judgments to random chance test score of 0.25

	N	Mean	Standard Deviation	95% Confidence Interval	T	P
Pre-Training Data One-Sample T: Qualitative Judgments	324	0.32	0.47	0.27, 0.37	2.73	0.003
Post-Training Data One-Sample T: Qualitative Judgments	236	0.29	0.45	0.23, 0.35	1.29	0.099

Each qualitative judgment was analyzed against the "Reference" standard to determine if bias was present. A histogram of all individual qualitative judgments was plotted showing the bias occurring across all of the judgments before and after training (Figure 2.7). The qualitative judgments made before training appeared to be biased low since ~52% of qualitative judgments were below the "Reference" category while only ~16% of qualitative judgments were above the "Reference" category. The rule-of-thumb training did not appear to shift the bias to slightly overestimate exposure as with the DIT score since ~47% and ~24% of qualitative judgments were below and above the "Reference" category, respectively.

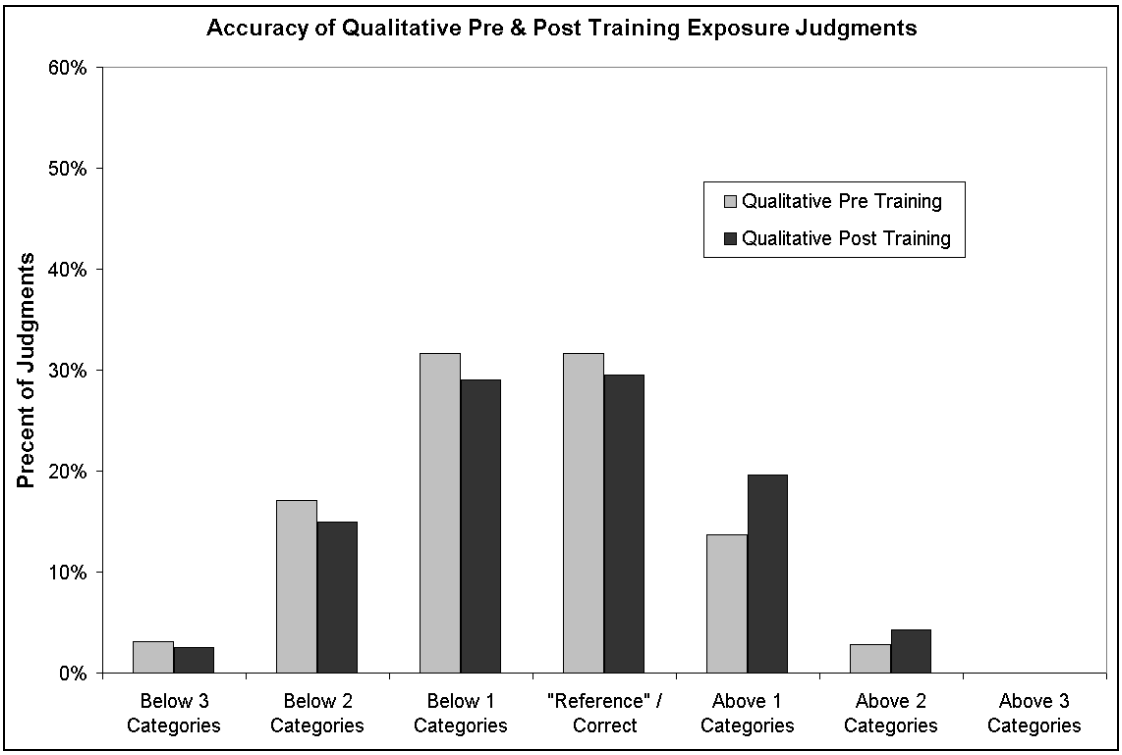

Figure 2.7: Percentage of all pre- and post-training qualitative task judgments above, below and reference categories

This study did not allow for the use of surrogate sampling data or facilitate the use of physical-chemical exposure models to help support making qualitative judgments. This type of information is often available for similar agents or jobs and provides good insight to the possible exposures in a given job or task. This was not a reasonable option for this desktop study since presenting surrogate data could lead to availability or anchoring and adjustment biases (Gilovich et al, 2002). Additional field studies could be initiated to better investigate the impact of accuracy when using surrogate or exposure modeling data to make qualitative judgments.

Since the rule-of-thumb training focused only on proper data interpretation of sample data, it is not surprising that the qualitative judgment accuracy did not improve post-training. To have a significant impact on accuracy and bias for qualitative judgments, attributes and rules that impact qualitative judgment accuracy would need to be identified. These attributes or determinants are more likely to be related to either experience that is specific to a selected task or the ability to utilize models for estimating potential exposures. It has been suggested that the use of physical-chemical models incorporating properties such as vapor pressure and scenario characteristics such as contaminant generation rate and ventilation rate can produce more accurate exposure assessments than limited sampling data (Jayjock, 1997, Nicas and Jayjock, 2002, Nicas, 2003). The framework used in this study could be easily modified to include determinants for

the use of physical-chemical models or environmental determinants to test the impact on qualitative judgment accuracy.

Quantitative Judgments and Training

The quantitative judgments were made following the initial judgment when sample data was presented one sample point at a time. This is probably similar to how sampling data is utilized during many routine field investigations. Hygienists typically cannot wait to get a statistically large number of samples before an exposure judgment is made. Judgments are more typically made or updated each time a new monitoring data point is received from the laboratory (Post et al, 1991, Ramachandran, 2008). Hygienists are required to make an exposure judgment and specify whether controls are needed based on the available basic characterization information and whatever sample data is available at the time. Many judgments made with only a few data points have a high level of uncertainty depending on the sample in relation to the exposure limit and other relevant surrogate data and information. This illustrates why Bayesian statistics and other probabilistic methods are attractive to hygienists who are forced to make judgments with sparse sampling data (Ramachandran, 2008).

As with the DIT score and qualitative judgments, task judgments with sampling data also had a random chance of being correct 0.25 or 25% of the time since only four exposure categories are available. The mean of all pre- and post-training quantitative scores for the group were 0.43 and 0.63, respectively (Table II.VI). The 95% lower bound for both the pre- and post-training was well above what would be expected from random chance. The upper bound of the mean before training was 0.47, which is below the lower bound of 0.58 after training. This indicates that the rule-of-thumb training provided a statistically significant positive impact to quantitative judgment accuracy. The percent of task judgments scoring above 50% correct went from 41% before training to 66% after training (Figure 2.8), illustrating the positive impact on quantitative judgment accuracy from the rule-of-thumb training.

Table II.VI: Summary of t-Tests comparing accuracy of pre-and post-training quantitative judgments to random chance test score of 0.25

	N	Mean	Standard Deviation	95% Confidence Interval	T	P
Pre-Training Data One-Sample T: Quantitative Judgments	324	0.43	0.42	0.38, 0.47	7.55	<0.001
Post-Training Data One-Sample T: Quantitative Judgments	236	0.63	0.38	0.58, 0.67	15.2	<0.001

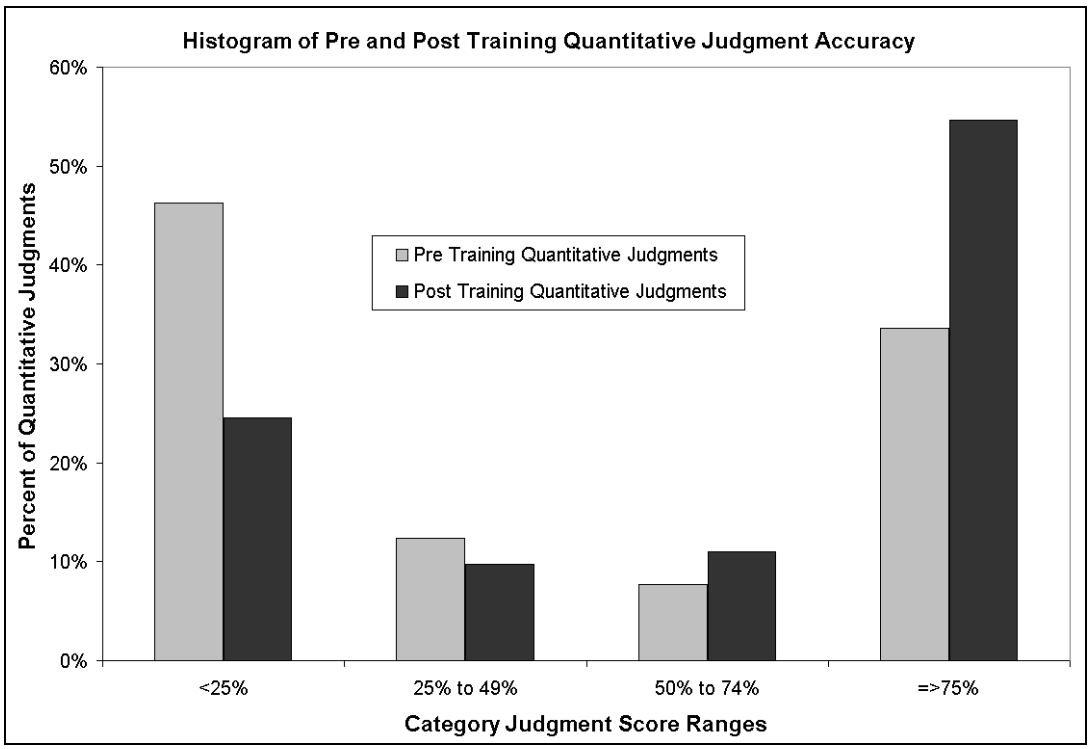

Figure 2.8: Histogram of pre- and post-training task quantitative judgment accuracy

Each quantitative judgment was also analyzed against the reference to illustrate potential bias. A histogram of all individual qualitative and quantitative judgments was plotted, showing the bias occurring across all of the judgments pre- and post-training (Figure 2.9). Judgments made before training appeared to be biased

low: ~47% of all pre-training judgments were below the reference category while only ~10% of all pre-training judgments were above the "Reference" category. The rule-of-thumb training appears to significantly reduce bias of post-training judgments: ~19% and ~19% were below and above the reference category, respectively.

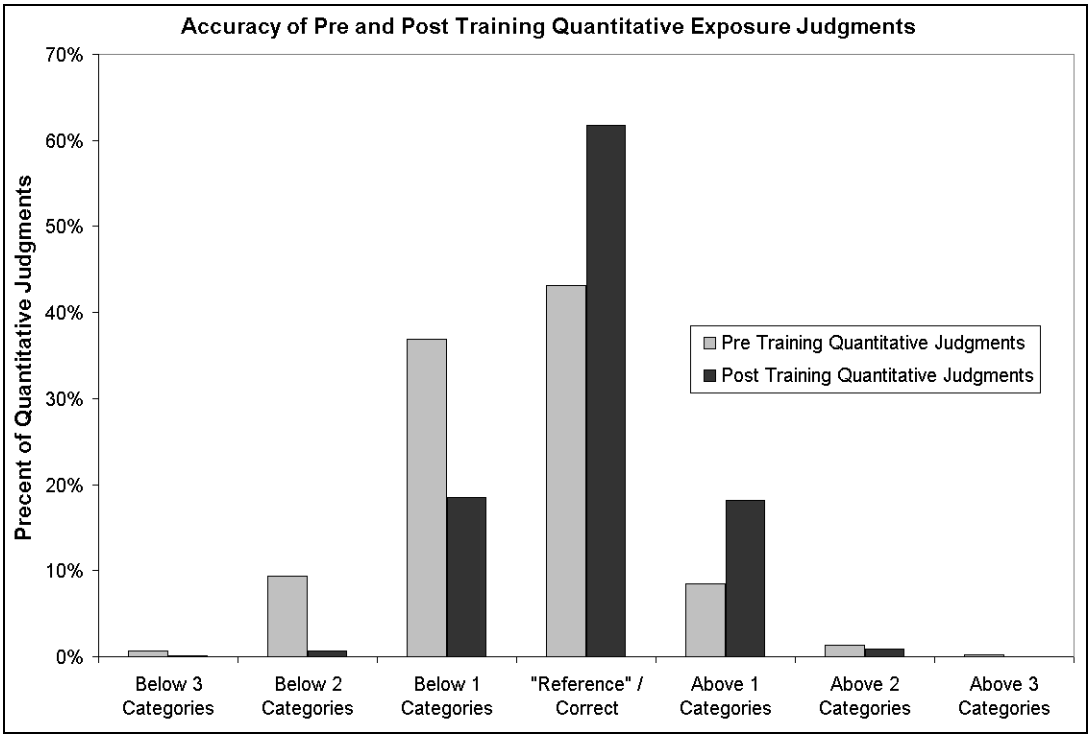

Figure 2.9: Percentage of all pre- and post-training quantitative task judgments above, below and reference categories

Evaluation of the DIT and quantitative judgment scores indicate that hygienists can make biased judgments when they do not use statistical tools. By inspection of specific judgment sets made by individual participants, it appears that some were anchored to the specific data point even after the rule-of-thumb training. Task 1 in Figure 2.10 illustrates how it was significant across the whole group judgments. Specifically, 80% of the post-training judgments were correct for the second sample in Task 1 for the group. The second sample given for Task 1 happened to be above the exposure limit. Therefore, understanding probability, the majority of participants selected the correct category, 4. However, the judgment accuracy for the third sample dropped to about 43%, indicating that many of the participants were *anchored* to the third sample, which was ~10% of the exposure limit. This was one of the more dramatic examples of collective bias (Surowiecki, 2005) due to anchoring and base-rate error found in the desktop study.

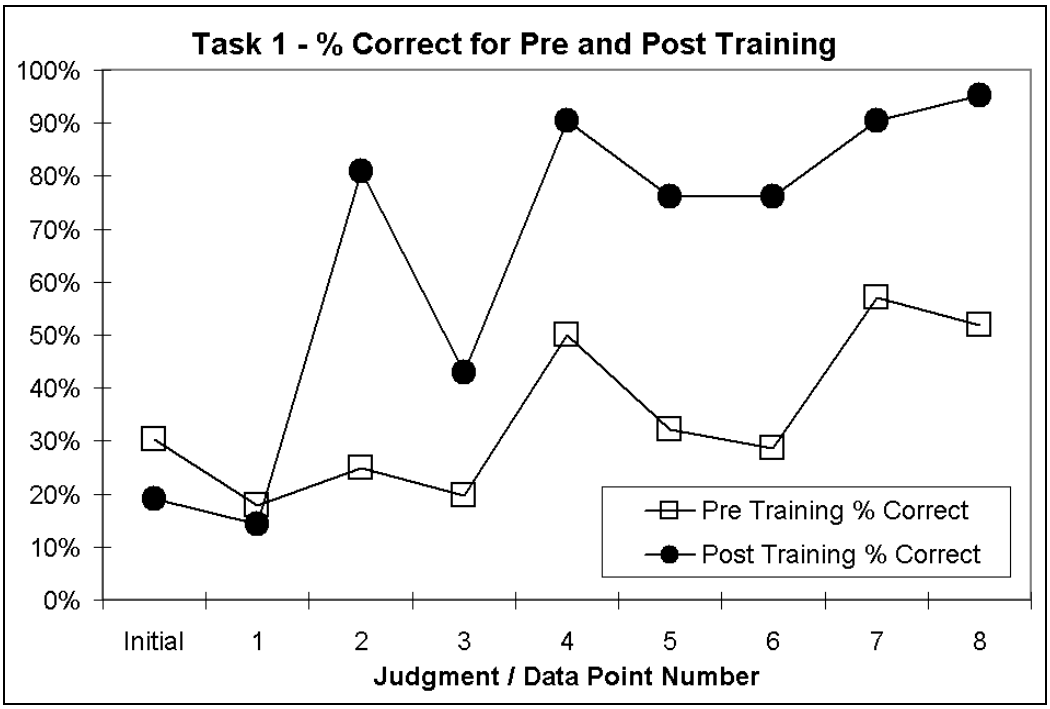

Figure 2.10: Charts of pre- and post-training judgment accuracy and overestimation for Task 1

CONCLUSIONS

These results highlight the importance of a basic understanding in lognormal parametric statistics and the routine use of statistical tools by occupational hygienists when making exposure judgments based on monitoring data. In this study, the use of simple lognormal rule-of-thumb training significantly improved the accuracy of participant's desktop quantitative exposure judgments and reduced bias of underestimating exposures. One can readily postulate that more consistent and broader use of robust statistical tools would significantly increase the accuracy of quantitative judgments and further reduce bias. Bayesian decision tools incorporated into the AIHA exposure assessment and management strategy provides a comprehensive and transparent framework for ensuring accurate exposure judgments. In addition, it appears that systematically documenting judgments and providing feedback on judgment accuracy can provide valuable training and calibration for hygienists, while helping identify specific ways to improve judgments (Plous, 1993, Gigerenzer, 1999, Meyer and Booker, 2001, Connolly et al, 2000, Hubbard, 2007).

Participant desktop qualitative judgments were little better than random chance and were not improved after statistical rule-of-thumb training. This is not surprising as the statistical rules-of-thumb are used in context with quantitative data. However, there is promise for improvement offered by the Bayesian integrated AIHA strategy as a framework that provides a powerful feedback mechanism for comparing qualitative judgments (Bayesian prior) to quantitative data interpretation (Bayesian likelihood). Further

research is needed to explore the effect of this important learning feedback loop on the accuracy of occupational hygienists' exposure judgments.

As with any study, care should be taken when interpreting findings. This study was performed on a relatively narrow number of tasks found in most industries. Therefore, the findings on qualitative judgment accuracy cannot be directly applied across all tasks and industries. The experience and education determinant distributions probably do not reflect the actual experience and education all of today's practicing hygienists. A study that would provide more insight about the accuracy of all qualitative judgments would need to include representative tasks across all industries and hygienists with education and experience determinants representative of the current hygiene profession.

Studies have been performed in various fields to investigate methods for identifying and controlling bias for faulty decision rules or heuristics (Kahneman et al, 1982, Gilovich et al, 2002). Generally, the mechanisms prescribed to control bias and increase accuracy focus on eliminating flawed decision rules by documenting the processes used to make judgments, creating feedback mechanisms and providing analytical tools and heuristics for experts (Gilovich et al, 2002, Hubbard 2007). Anecdotal evidence suggests that industrial hygienists may not systematically use statistical tools while making decisions based on small data sets (Ramachandran, 2008). This observation may be due to either a general discomfort with statistics or the limited number of statistical tools that can effectively analyze small data sets available to hygienists. Statistical tools that allow for small data sets and provide output in terms of probability could be useful in increasing their usage across the profession. Specific training can then be designed and implemented to focus on the specific elements that improve judgments. If there are several heuristics that are quite common in making judgments across many professions, it is likely that they are also used when making exposure judgments. The insights of heuristics and biases identified in many other studies and professions can help occupational hygienists identify tools and strategies that increase exposure judgment accuracy and reduce bias. This, in turn, will help occupational hygienists utilize available resources to better understand potential exposures and protect workers.

Subsequent discussions on selection of a decision statistic generated valuable discussion and may help to calibrate hygienists when interpreting exposure data as well as provide a "frequentist" format for the decision statistic (Gigerenzer and Hoffrage, 1995, 1999). Periodic use of data interpretation tests (DITs) can help calibrate hygienists and will likely increase quantitative judgment accuracy. The rule-of-thumb training was a simple exercise that provided significant accuracy improvements to data interpretation exercises and quantitative task judgments. In addition, other simple and well designed data interpretation rules may also increase judgment accuracy. Even though the rule-of-thumb training provided significant

improvements in accuracy, it is more important that statistical tools be used consistently when interpreting sampling data.

The framework presented in this study can be modified and applied to many other investigations of exposure judgment determinants, bias and accuracy. Analysis of DITs and exposure judgments in further studies may help better understand possible sources of biases from heuristics used by hygienists. Understanding the impact of the exposure judgment elicitation exercises and data interpretation training on exposure judgment accuracy in this desktop study can further refine exposure judgment studies performed in the field. The implementation of a rule of thumb or other heuristic training and analysis of other determinants that effect accuracy can help identify ways to continuously improve judgment efficiency and accuracy. It is hoped that the findings in following professional judgment studies will provide hygienists with tools to more efficiently and effectively manage potential exposure, providing benefits to workers and companies alike. This especially includes field studies where physical-chemical modeling and other factors affecting qualitative exposure judgment accuracy could be analyzed and published for the benefit of workers, companies and the occupational hygiene profession.

CHAPTER 3 - Desktop study of occupational exposure judgments: Does education and experience influence accuracy?

Accurate exposure assessments provide a foundation for effective exposure management programs that protect employees and communities. This study examines the impact of several experience and education determinants on exposure judgment accuracy. The study used desktop assessments performed on several different jobs with different exposure profiles to identify correlations between determinates and judgment accuracy using logistic regression models. The judgments for each chemical on each different task were elicited from industrial hygienists with varying levels of experience, education and training. Videos, written and oral information about the exposure tasks were presented to all participants as they documented a series of qualitative and quantitative exposure judgment probabilities in four exposure categories while sampling data was presented serially. Participants (n=77) first documented their qualitative exposure assessment and then their quantitative exposure assessments after receiving the series of sampling data points. Data interpretation tests and training in simple rule-of-thumb data interpretation were also given to each participant to investigate the impact of data interpretation skills on exposure judgment accuracy. Logistic regression analysis indicated "years of exposure assessment experience" ($p<0.05$), "highest EHS degree" ($p<0.05$) and a participant's "data interpretation test score" ($p<0.05$) directly impacted qualitative exposure judgment accuracy. Logistic regression models of quantitative judgment accuracy showed positive correlation with "greater than 10 years of exposure assessment experience" ($p<0.05$), "highest EHS degree" ($p<0.05$), a participant's "data interpretation test score" ($p<0.001$), rule-of-thumb data interpretation training ($p<0.001$), and the number of sample data points available for a judgment ($p<0.005$). Analyzing judgments in subsets for participants with fewer or more than 10 years experience indicated additional correlations with Certified Industrial Hygienist and Certified Safety Professional certifications, total number of task exposure assessments, and career number of air surveys. The correlation of qualitative and quantitative exposure judgment accuracy with "greater than 10 years experience" supports similar research findings from other fields. The results of this study indicate that several determinants of experience, education and training, in addition to the availability of sampling data, significantly impact the accuracy of exposure assessments for the set of exposure tasks and agents used in this study. The findings also suggest methods for enhancing exposure judgment accuracy through statistical tools and specific training.

INTRODUCTION

Exposure assessments provide the foundation for determining whether occupational and environmental exposure risks are efficiently and effectively managed (Ignacio, 2009). Occupational hygienists utilize

available information on agents, workforce, equipment, materials, work practices, existing exposure controls and their professional judgment to quantify the exposure risk by arriving at an exposure assessment or judgment. Exposure judgments are commonly used in a wide range of situations, including retrospective exposure assessments for epidemiology studies, and current and prospective exposure assessments for managing exposures related to consumer use and manufacturing operations(Hawkins, 1989, Teschke, 1989, Macaluso, 1993, Friesen, 2003, Ramachandran, 2003). When there are limited sampling data, occupational hygienists routinely use professional judgment to assess the acceptability of exposures for managing engineering controls, medical surveillance, hazard communication and personal protective equipment programs (Ignacio, 2009). Professional judgment in the context of exposure assessment can be considered "the capacity of an experienced professional to draw correct inferences from incomplete data" based on their "knowledge gained from formal education, experience, experimentation, inferences and analogy" (Ignacio, 2009). However, the accuracy of the exposure assessment's underlying decisions and performance of a given exposure assessment strategy used are not often studied formally and published (Ramachandran, 2008).

A considerable number of diverse chemical agents can exist in occupational environments where workers may have a wide distribution of exposure over time. Much of this variability can be introduced due to a diverse number of exposure tasks, agents and work practices (Kromhout, 1987). This complexity highlights the need for a well-designed comprehensive exposure assessment strategy and effective tools to efficiently arrive at accurate judgments for managing exposure risks. The American Industrial Hygiene Association's (AIHA) "A Strategy for Assessing and Managing Occupational Exposures" is a well-known strategy that provides a simple yet elegant framework used for performing and managing exposure assessments (Ignacio, 2009). The framework utilizes an efficient method to select exposure control categories defined by fractions or multiples of a selected agent's health-based or compliance-based exposure limit to classify a given exposure judgment (Table III.I). These exposure control categories in the AIHA strategy also facilitates a transparent and effective "control banding" method for agents which formal exposure limits have not been developed. Qualitative and quantitative exposure assessments are performed after a thorough review of available information and data related to the workforce, jobs, materials, worker interviews, exposure agents, exposure limits, work practices, engineering controls and protective equipment. A comprehensive and flexible strategy facilitates refinement of exposure assessments when new information or data becomes available (Ramachandran, 2008).

Table III.I: AIHA Exposure Category Rating Scheme. A SEG is assigned an exposure rating by comparing the 95th percentile exposure distribution ($X_{0.95}$) with the full-shift time-weighted average (TWA), Occupational Exposure Limit (OEL) or Short-Term Exposure Limit (STEL) to determine in which category it most likely falls (Ignacio, 2009).

AIHA Exposure Rating	Proposed Control Zone Description	General Description	AIHA-Recommended Statistical Interpretation
1	Highly Controlled (HC)	95th percentile of exposures rarely exceeds 10% of the limit.	$X_{0.95} \leq 0.10$ OEL
2	Well Controlled (WC)	95th percentile of exposures rarely exceeds 50% of the limit.	0.10 OEL $< X_{0.95} \leq 0.5$ OEL
3	Controlled (C)	95th percentile of exposures rarely exceeds the limit.	0.5 OEL $< X_{0.95} \leq$ OEL
4	Poorly Controlled (PC)	95th percentile of exposures exceeds the limit.	OEL $< X_{0.95}$

Studies suggest that exposure judgments are routinely made using "professional judgment" to arrive at a qualitative exposure assessment when sampling data is not available by using available information coupled with the assessor's experience (Hawkins, 1989, Teschke, 1989, Macaluso, 1993, Friesen, 2003, Ramachandran, 2003, Kromhout, 1987, Rock, 1986, Logan, 2009). Referring to published research, one may conclude that qualitative exposure judgments are difficult, highly variable and require specific "experience" to be accurate. Several recent publications have suggested using a categorical-based exposure assessment framework to facilitate understanding experience, education and training, which effect exposure assessment accuracy (Logan, 2009, Hewett, 2006). Quantitative exposure judgments using sampling data would also require relevant experience and specifically the experience with using statistical concepts and tools when interpreting the data. "Experience" can be specific to types of operations, jobs, tasks or chemicals, and it can also be general, based on total time or total number of assessments, not specific to a given task or chemical. Professional certifications, such as the Certified Industrial Hygienist or CIH (ABIH, 2010) Certified Safety Professional or CSP (BCSP, 2010) or other related certifications could be considered a broad combination of experience, training and education due to their comprehensive scope. Intuitively, most would expect that the more training and education an exposure assessor accumulates, the more accurate their judgments will become. In general, various links between training and specific education with performance has been well established in other fields of study (Goldstein, 1997, Ericsson, 1993).

Recent work by Logan used the AIHA strategy for making exposure assessments to indicate that specific training and sampling data interpretation skills can significantly increase a hygienist's ability to make accurate quantitative exposure judgments (Logan, 2009). The study discussed possible heuristics or mechanisms that may be used for performing exposure judgments and the biases that could result from their use. In addition, the authors discussed methods for specific data interpretation training to further increase quantitative exposure judgment accuracy based on study observations and literature reviews from other scientific research (Goldstein, 1997, Ericsson, 1993, Gilovich, 2002, Montgomery, 2005, Zsambok, 1997, Gigerenzer, 1999, Klein, 1998) Logan et al introduced simple data interpretation training that was specifically designed for use with limited sample data sets and the AIHA strategy (Logan, 2009, Appendix I). The training, administered to participants in their study, significantly impacted the accuracy of quantitative exposure judgments made without the use of statistical tools. However, the specific data interpretation training did not increase qualitative exposure assessment accuracy. In summary, the study recommended further work to better understand the relationships between various types of experience, education or training with exposure assessment accuracy. This study aims to begin investigating a larger set of experience, training and education determinants for significant correlations with qualitative and quantitative exposure assessment accuracy.

BACKGROUND

Over the past several decades, investigators across various fields have studied factors that contribute to high performance and identified common mechanisms that contribute to poor performance. [(14-18)] Improvements in performance associated with experience and well-designed training appear to occur in a wide variety of professions, including physicians, athletes, military officers, musicians, lawyers, engineers, chess players and students (Goldstein, 1997, Ericsson, 1993, Gilovich, 2002, Montgomery, 2005, Zsambok, 1997, Gigerenzer, 1999, Klein, 1998). These studies suggest updating the old saying 'practice makes perfect' to '*well designed* practice makes perfect' meaning that education and training should be carefully designed to enhance the specific skills that most drive overall performance. This probably seems intuitive to anyone familiar with training routines of professional athletes or musicians. However, it may not seem obvious for complex cognitive skills based on education, training and experience used in making a medical diagnosis or an exposure judgment.

Decisions or judgments are made using a complex blend of cognitive rules and beliefs combined with emotional and unconscious mechanisms (Gilovich, 2002, Kahneman, 2000, Ericsson, 2006). Biases resulting from the unconscious elements of common heuristics can significantly impact the accuracy of various judgments (Goldstein, 1997, Ericsson, 1993, Gilovich, 2002, Montgomery, 2005, Zsambok, 1997, Gigerenzer). The combinations of emotional or unconscious mechanisms are sometimes categorized as

intuition or heuristics and can be influenced and even calibrated by deliberate practice (Ericsson, 1993, Gilovich, 2002, Montgomery, 2005). This finding appears to hold across many areas of science, including exposure assessment. A handful of studies related to accuracy or agreement of occupational and environmental exposure judgment have been published over the past few decades (Hawkins, 1989, Teschke, 1989, Macaluso, 1993, Friesen, 2003, Ramachandran, 2003, Ramachandran, 2008, Kromhout, 1987). In general, these studies indicate that judgments improve with more information and data, well designed strategies can increase effectiveness and efficiency, and professional judgment is often better than random chance. A few studies suggest that the common cognitive biases found in many fields have the potential to impact the accuracy of exposure judgments as well (Logan, 2009, Walker, 2001, Walker, 2003). To identify methods and mechanisms for improvement, research recommends the need for further study of determinants and heuristics impacting judgment accuracy. These determinants can be defined as representing various elements of education, training and experience. Education can be defined formally, such as degrees obtained from a university or more specific education domains obtained in particular professional development courses available through conferences and associations. Training can be represented in general or very specific areas and targeted to address specific biases that are common to many people. Experience can be defined broadly in terms of time or specifically in terms of numbers of activities related to the outcome being studied (Meyer, 2001, Hubbard, 2007). As training, education and experience determinants of judgment accuracy are better understood, studies can be designed to better understand underlying judgment algorithms or heuristics and help define effective training and education that enhances professional judgment performance.

It is also important to consider that for complex, cognitive tasks, conscious decision-making algorithms are usually more accurate but typically slower than intuitive or emotional decision making (Gilovich, 2002). Researchers find that these complex decision-making processes can become much faster through prolonged, proper training and ultimately lead to better performance (Griffin, 1992, Gigerenzer, 1999, Connolly, 2000). Studies of professional athletes have reported that specifically designed activities contribute more to elite status than general experience within the domain of a particular sport or activity (Ericsson, 2006). Proper experience is needed to achieve “expert” status regardless of talent in a wide variety of fields. This supports the notion that specialized skills require continuous maintenance and can take many years to develop. Specifically, research suggests that it takes 10 or more years to accumulate enough hours of “deliberate practice” for mastering a particular domain (Ericsson, 1993). Intuitive processes or heuristics are much more efficient than cognitive processes, and through many hours of training most professionals or experts have developed a broader set of heuristics specific to their domain of practice. As a result, experts have been shown to make faster and more accurate decisions than novices. This development of effective heuristics is most often achieved by well-designed “deliberate practice” of cognitive, conscious decision-making skills that evolve into heuristics or intuitive decision-making

mechanisms. Though there are variations in what is considered deliberate practice, in general it is made up of: (1) well-designed education and training plans with clear goals; (2) repeated practice experiences designed to improve core skills and functions; and (3) feedback mechanisms measuring the quality of each attempt and overall performance. There appears to be significant research published supporting the idea that expertise in any domain is mostly driven by deliberate practice regardless of field (Gilovich, 2002, Ericsson, 2006). Since occupational hygiene judgment can require many sources of information and complex logic, it is expected that the concept of increased performance through deliberate practice also applies to making accurate and efficient exposure judgments.

METHODS

The judgments in this study were elicited from self selected groups of industrial hygienists who attended sessions where informed consent was obtained (n=77). The hygienists had varying levels of experience, education and related determinants for a group of common tasks found in industry. All determinants were documented on paper by each participant prior to exposure judgment elicitations. Videos, written and oral information was presented to all participants for each scenario as they documented exposure judgment probabilities in the AIHA exposure control categories while sampling data was presented in a serial manner. Each of the tasks had sampling data sets ranging from 6 to 10 samples. Data interpretation tests (DITs) and training in simple rule-of-thumb exposure data interpretation was also given to investigate the impact of data interpretation skills on exposure judgment accuracy. The DITs were administered prior to collecting exposure judgments on videotaped tasks and directly after the simple *"Rule-of-thumb"* data interpretation training was given. A *"DIT Score"* was calculated for each participant's data interpretation test using equation 1.

"DIT Score" = (number correct judgments) / (8 datasets) (Equation 3.1)

Detailed descriptions of a DIT and the simple rule-of-thumb were previously published by Logan et al. [(10)] Summary statistics for each experience, education and related determinants are listed in Tables II and III for all participants in the study. Table II lists general education and experience descriptions and Table III lists experience specific to agents or jobs selected in the study.

The education and experience determinants could only have a few possible responses in the given categories. For analysis, all of the education and experience determinant variables are considered binary or ordinal since the possible categories for the determinants ranged from 2 to 5 possible selections. The qualitative and quantitative judgments were constructed into binary data described by a 0 or 1 for incorrect and correct judgments respectively. In order to analyze the effect of the education and experience

determinants on the exposure judgments, binary logistic regression modeling methods were used. Analysis of all determinants was done using a binary logistic regression (Pagano, 1993) function on qualitative and quantitative exposure judgment variables using Minitab® software version 15 (Minitab Inc., State College, PA 16801). An analysis of correlation was performed between various combinations of determinant variables to support the most relevant selection of model variables by calculating the Pearson product moment correlation coefficients. To strengthen the overall regression model and applicability of results to general practice, selections of a variable between two highly correlated variables were made based on which variable was most relevant for interpretation. If two highly correlated variables represented different aspects of education or experience, an interaction term from both variables was created to construct more accurate and relevant logistic regression models. An interaction variable denoted ***"CIH&CSP"*** for the CIH and CSP certification determinants was created to account for significant correlation between these variables throughout the data set. Significant correlation between "***Rule-of-Thumb***" training and ***"DIT Score"*** was accounted for using an additional interaction variable in each model**.**

TABLE III.II – Participant-specific determinant variables and descriptions of selections

Participant-specific Determinant Variable	Levels of Determinant Variable	Count	Percent
Highest EHS or Science Degree Achieved	1 = Associates degree or no degree	10	13%
	2 = Undergraduate degree	19	24%
	3 = Master's degree	46	60%
"High EHS Degree"	4 = Advanced degree (PhD, ScD, or other)	2	3%
CIH Certification	0 = No	29	38%
"CIH"	1 = Yes	48	62%
CSP Certification	0 = No	55	71%
"CSP"	1 = Yes	22	29%
CIH and CSP Certifications	0 = No	58	75%
"CIH&CSP"	1 = Yes	19	25%
Number of years making exposure judgments	1 = <1	12	16%
	2 = 1-3	6	8%
	3 = 3-5	10	13%
"Years EA"	4 = 5-10	15	19%
	5 = >10	34	44%
Estimate of the Total Number of Job-task Exposure Assessments Documented	1 = <10	23	30%
	2 = 11 to 100	14	18%
	3 = 101 to 1000	27	35%

Using AIHA Strategy ***"AIHA EA Cat"***	4 = 1001 to 5000	8	10%
	5 = > 5000	5	6%
Experience in Mathematical Modeling of Exposures ***"ModelExp"***	1 = Never used a mathematical exposure model	17	22%
	2 = Used only simple exposure model (i.e., saturation vapor pressure) a few times	29	38%
	3 = Some experience using different exposure models	29	38%
	4 = Frequently use exposure models	1	1%
	5 = Have designed mathematical exposure models	1	1%
Experience in Exposure Assessment-related Statistics ***"StatsExp"***	1 = Do not have a good understanding of normal and lognormal statistics	10	13%
	2 = Can do basic lognormal statistical calculations (GM, GSD)	37	48%
	3 = Familiar with calculating point estimate of arithmetic mean & 95%tile	11	14%
	4 = Familiar with calculating exceedance fractions and tolerance limits	19	25%
Career Air Sampling Survey Experience ***"CareerAirSurveys"***	1 = <3	7	9%
	2 = 4 to 20	9	12%
	3 = 21 to 100	11	14%
	4 = 101 to 200	19	25%
	5 = >200	31	40%

TABLE III.III – Task- or chemical-specific experience determinant variables for all participants on each of the tasks they performed judgments.

Task- or Chemical-specific Determinant Variable	Levels of Determinant Variable	Count	Percent
Years Experience with Exposure Assessments on This Type of Task[A]	1 = Never	225	42%
	2 = 1 to 2 years	127	24%
	3 = 3 to 5 years	57	11%
"TaskYearExp"	4 = 6 to 10 years	60	11%
	5 = >10 years	61	12%
Approximate Number of Exposure Assessments on This Chemical[B]	1 = 0	275	52%
	2 = 1 to 2	84	16%
	3 = 3 to 5	51	10%
"TaskNumEAs"	4 = 6 to 10	47	9%
	5 = >10	73	14%
Experience in Sampling Specific Chemical Agent[C]	1 = never sampled it	302	57%
	2 = observed sampling	20	4%
	3 = collected 1 to 2 surveys on agent	79	15%
"NumAirSurveysChem"	4 = collected 3 to 10 surveys on agent	65	12%
	5 = collected >10 surveys on agent	63	12%

[A] – This determinant includes a participant's experience with a given task and how many task judgments they made in each years of experience category. Each participant made judgments across multiple tasks (n=530).

[B] - This determinant includes a participant's experience making exposure assessments on a specific chemical into each years of experience category. Each participant had different experience for the different chemical agents (n=530).

[C] - This determinant includes a participant's experience conducting air sampling surveys for a specific chemical on a given task. Each participant had different sampling experience for the different chemical agents (n=529).

RESULTS

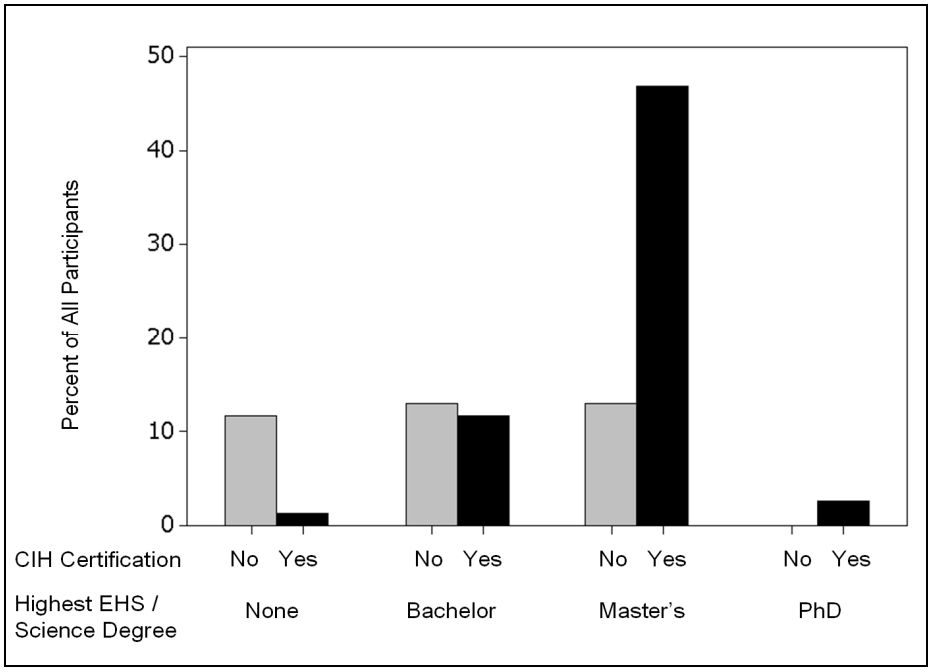

Figure 3.1- Percent of participants with CIH certification for each education level category

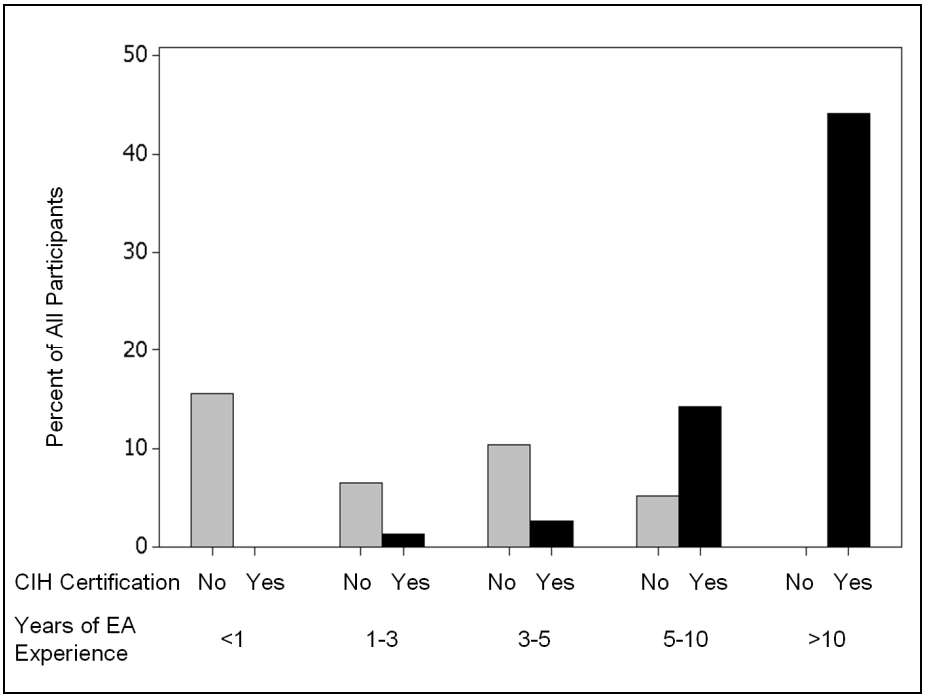

Figure 3.2 – Percent of participants with CIH certification for each years of exposure assessment experience category

More than half of participants had a Master's degree, almost a quarter of participants had a Bachelor's degree, and only 2 participants had a doctorate degree (Figure 1). Almost half of the total participants had both a master's degree in a science-related field and had a CIH certification. Approximately 44% of participants had more than 10 years exposure assessment experience and also held a CIH certification. Participants with increasing years of experience were also more likely to have a CIH certification (Figure 3.2). All participants with more than 10 years exposure assessment experience held a CIH certification, while less than 10% of participants with less than 5 years exposure assessment experience held a CIH certification.

Qualitative Judgments

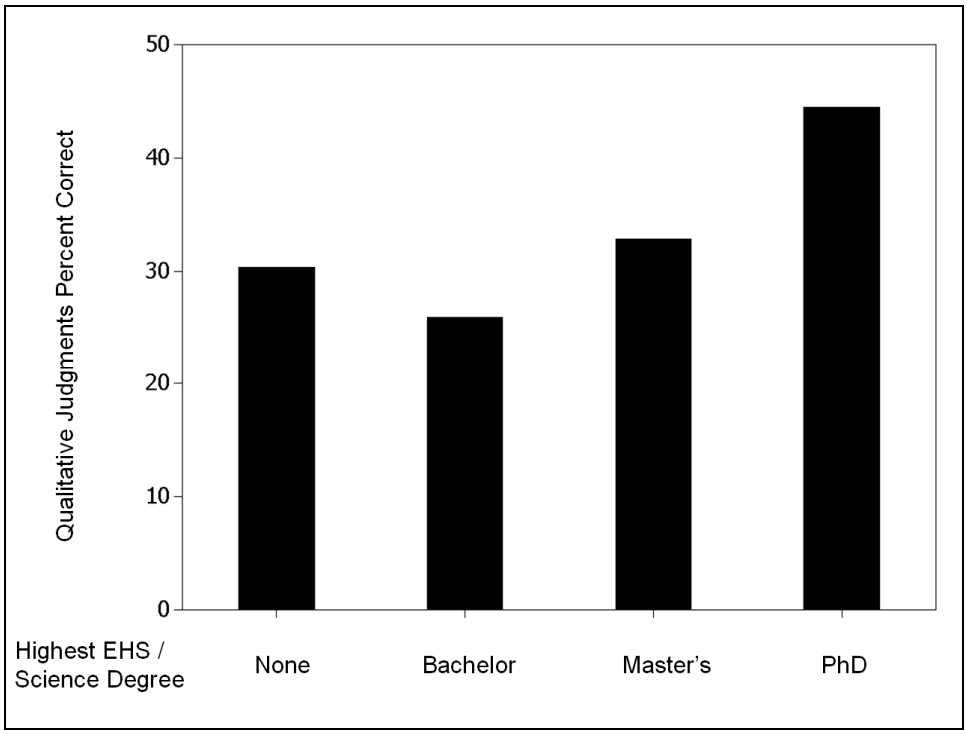

Figure 3.3 - Qualitative exposure judgments accuracy as a percentage within each "High EHS Degree" for all participants

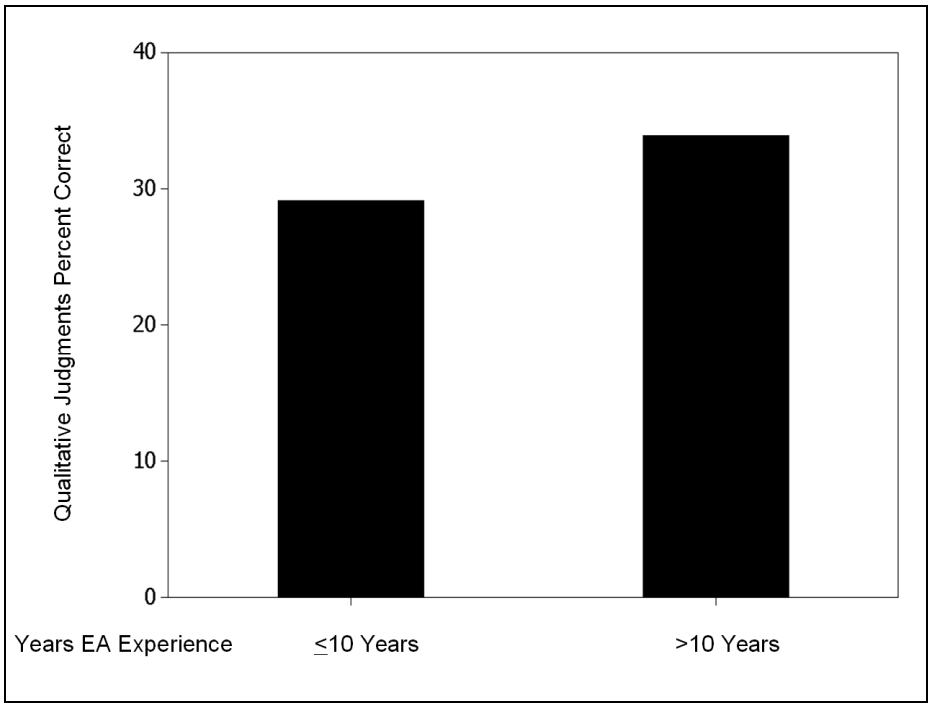

Figure 3.4 - Qualitative exposure judgments percent correct for ≤10 and >10 years of exposure assessment experience for all participants

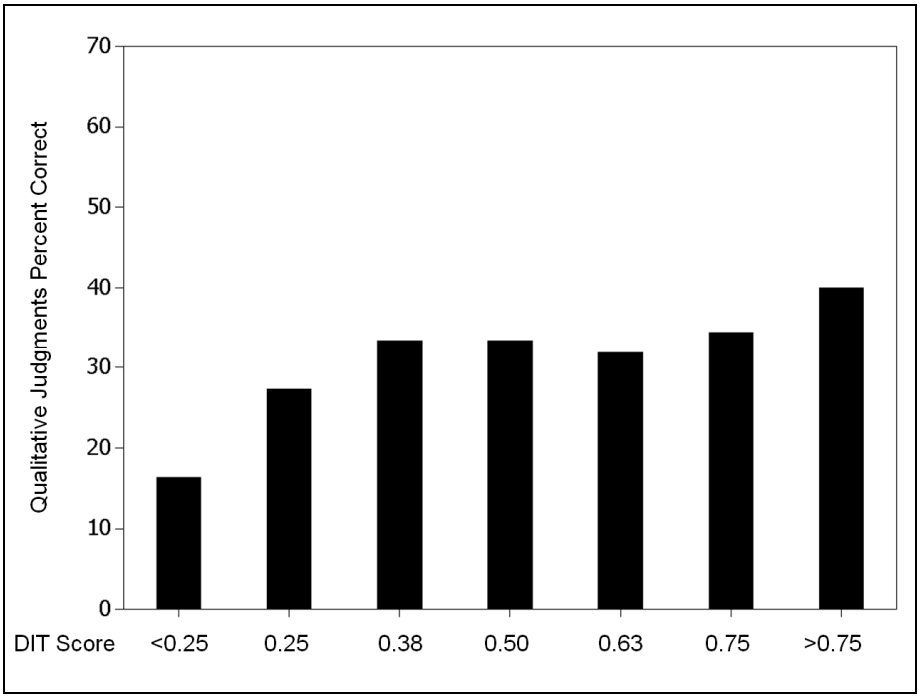

Figure 3.5 – Histogram of qualitative exposure judgments percent correct for each Data Interpretation Test Score or "DIT Score"

Qualitative judgment accuracy appears to increase with level of education shown in Figure 3.3, while a slight increase in judgment accuracy can also been seen for participants with greater than 10 years exposure assessment experience in Figure 3.4. Figure 5 shows that qualitative exposure judgment accuracy generally increased with increasing *"DIT Score"*. Each Data Interpretation Test had 8 datasets which allows for 9 possible discrete values of *"DIT Score"* ranging from 0.00 to 1.00.

TABLE III.IV – Binary logistic regression model results for all qualitative exposure judgments for education and experience determinants which were statistically significant (p<0.05). Analysis #1 used a categorical "Years EA" *variable for experience, while Analysis #2 used a binary variable of* "10 Years EA" *in the logistic regression model.*

	Analysis #1: All Qualitative Judgments using *"Years EA"*		Analysis #2: All Qualitative Judgments using *"10 Years EA"*	
	Z value	p value	Z value	p value
"High EHS Degree"	(A)	>0.05	2.03	0.043
"Years EA"	2.21	0.027	(B)	(B)
"10 Years EA"	(B)	(B)	(A)	>0.05
"DIT Score"	2.10	0.036	2.10	0.036

(A) - Variable was not statistically significant in the logistic regression model (p>0.05).

(B) - Not analyzed since *"Years EA"* and *"10 Years EA"* are categorical and binary representations of the same variable which cannot be included in the same regression analysis.

TABLE III.V – Binary logistic regression output of qualitative exposure judgments made by participants with fewer than 10 years' exposure assessment experience for education and experience determinants that were statistically significant (p<0.05) EA experience.

	Analysis #3 Qualitative Judgments for Participant's with <10 Years EA Experience		Analysis #4 Qualitative Judgments for Participant's with >10 Years EA Experience	
	Z value	p value	Z value	p value
"High EHS Degree"	2.15	0.031	(A)	>0.05
"TaskNumEAs"	2.42	0.015	(A)	>0.05
"DIT Score"	2.16	0.03	(A)	>0.05

(A) - Variable was not statistically significant in the logistic regression model (p>0.05).

Table III.IV and III.V represent the statistically significant determinants ($p<0.05$) from the logistic regression models for analysis of qualitative exposure judgments. Years of exposure assessment experience or "*Years EA*" (Tables III.IV and III.V) is a categorical variable included in Analysis #1 while the 10 or more years of EA experience variable is the binary representation for the "*Years EA*" (Analysis #2, Table III.IV). In both models, the continuous variable *"DIT Score"* was found to correlate with accuracy of qualitative exposure assessment, along with years exposure assessment experience (Analysis #1, Table III.IV). Data interpretation or rule-of-thumb training was not statistically significant in predicting qualitative judgment accuracy for any of the models. Logistic regression of participants with less than 10 years experience in Analysis #3 did show a significant correlation between "*High EHS Degree*", *"TaskNumEAs"* and *"DIT Score"* for qualitative exposure assessment accuracy (Analysis #3, Table III.V). None of the education and experience determinants were significantly correlated with qualitative judgment accuracy for participants with more than 10 years of exposure assessment experience (Analysis #4, Table III.V).

Quantitative Judgments

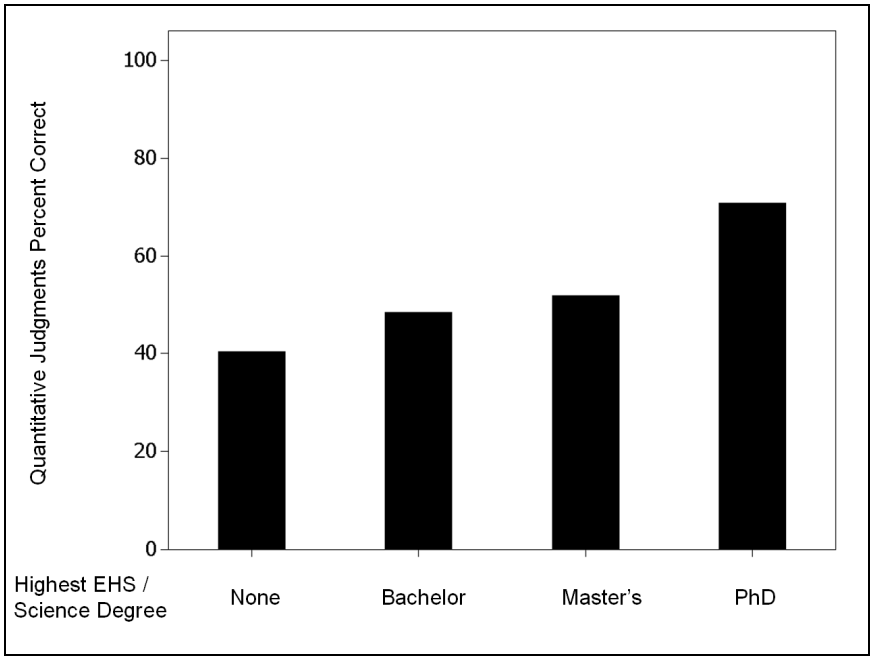

Figure 3.6 - Quantitative exposure judgments percent correct for each "High EHS Degree" education determinant

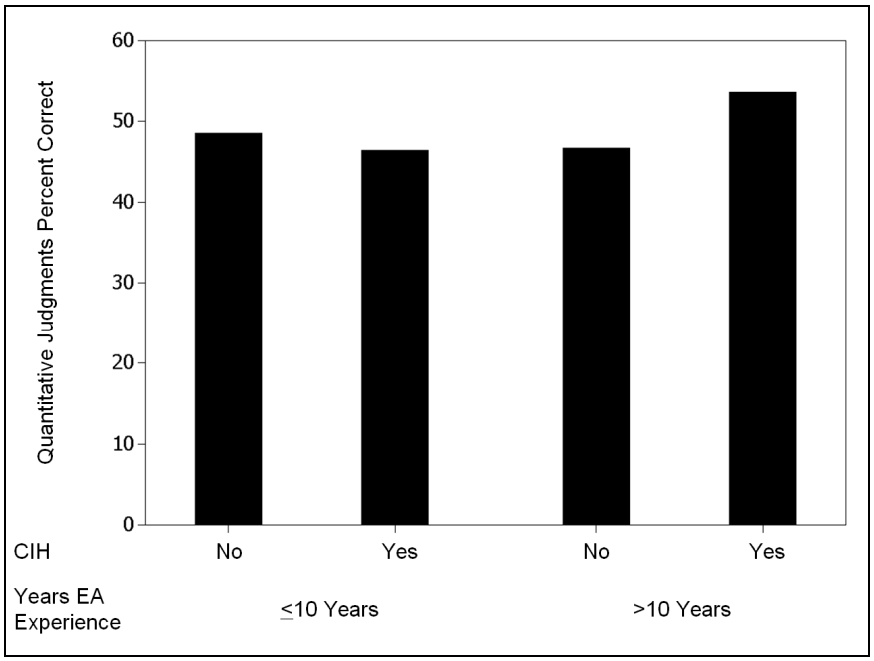

Figure 3.7 - Quantitative exposure judgments percent correct within each category for CIH and Years of Exposure Assessment experience < and > 10 years (n=3,772)

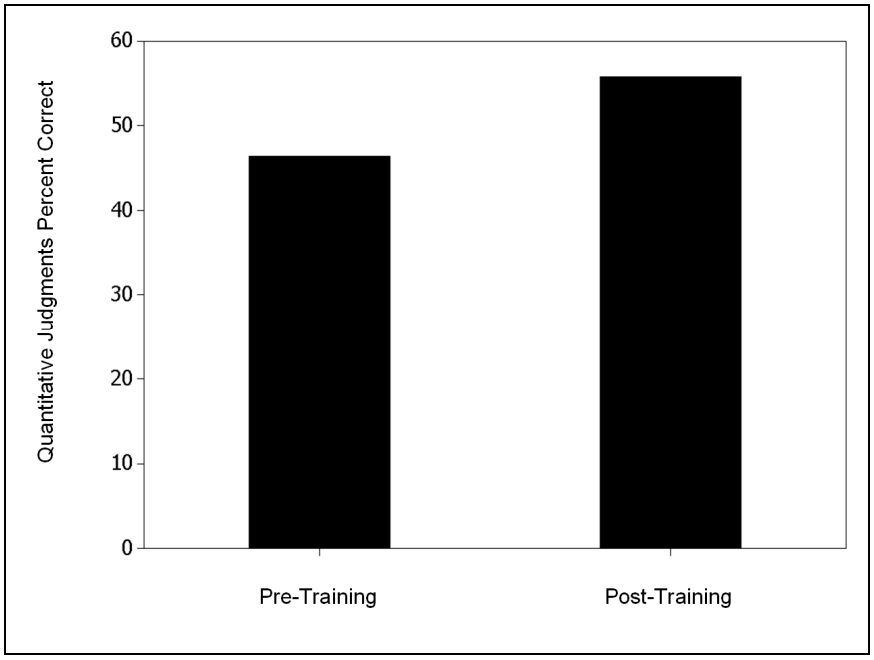

Figure 3.8 - Quantitative exposure judgments percent correct for pre and post data interpretation "Rule-of-Thumb" training

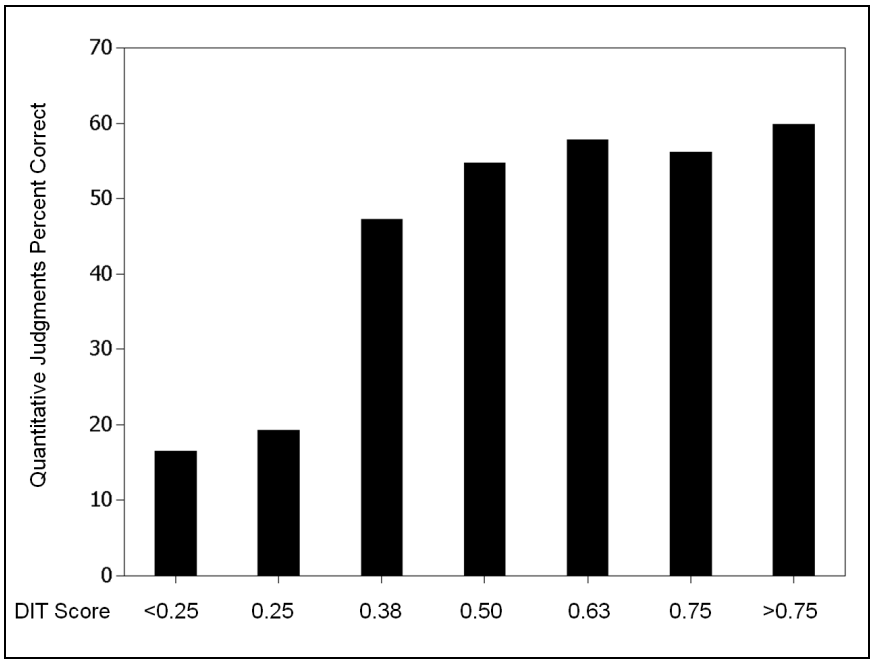

Figure 3.9 - Quantitative exposure judgments percent correct for participants across each possible "DIT Score"

Accuracy of all quantitative exposure judgments generally appears to increase across each level of degree as shown in Figure 3.6. Exposure assessors with both a CIH certification and >10 years of experience had >50% accuracy, while the other combinations of experience and CIH certification were <50% in each group (Figure 3.7). Quantitative exposure assessments made after the data interpretation or rule-of-thumb training were, as a whole, >50% accurate, while judgments made before training were <50% accurate (Figure 3.8). Quantitative exposure assessment accuracy tended to increase with increasing *"DIT Score"* (Figure 3.9), which was also seen with qualitative judgments.

TABLE III.VI – Binary logistic regression model results for quantitative exposure judgments with education and experience determinants that were statistically significant (p<0.05). Analysis #5 used a categorical "Years EA" *variable for experience, while Analysis #6 used a binary variable of* "10 Years EA" *in the logistic regression model.*

	Analysis #5 All Quantitative Judgments using *"Years EA"*		Analysis #6 All Quantitative Judgments using *"10 Years EA"*	
	Z value	p value	Z value	p value
"High EHS Degree"	2.12	0.034	2.24	0.025

"Years EA"	[A]	>0.05	[B]	[B]
"10 Years EA"	[B]	[B]	2.68	0.007
"DP" (number of sampling data points)	2.90	0.004	2.90	0.004
"DIT Score"	13.01	<0.001	12.88	<0.001
"Rule-of-thumb" Training	6.47	<0.001	6. 46	<0.001
"DIT Score" and *"Rule-of-thumb"* Interaction Variable	-7.02	<0.001	-7.08	<0.001

[A] - Variable was not statistically significant in the logistic regression model (p>0.05).

[B] - Not analyzed since *"Years EA"* and *"10 Years EA"* are categorical and binary representations of the same variable and therefore cannot be included in the same analysis. Analysis #5 used *"Years EA"* and *"CIH"* in the binary logistic regression model and neither were found to be significant.

TABLE III.VII –Binary logistic regression model results for quantitative exposure judgments with education and experience determinants that were statistically significant (p<0.05) for participants with less than or greater than 10 years EA experience.

	Analysis #7 Quantitative Judgments Participant's EA Experience <10 Years		Analysis #8 Quantitative Judgments Participant's EA Experience >10 Years	
	Z value	p value	Z value	p value
"CIH"	-4.04	<0.001	[B]	[B]
"CSP"	-2.01	0.045	[C]	[C]
"CIH&CSP"	3.80	<0.001	-2.53	0.011
"High EHS Degree"	2.01	0.045	4.55	<0.001
"CareerAirSurveys"	-2.26	0.024	[A]	>0.05
"DP" (number of sampling data points)	2.42	0.016	[A]	>0.05
"DIT Score"	9.67	<0.001	7.88	<0.001
"Rule-of-thumb" Training	6.16	<0.001	[A]	>0.05
"DIT Score" and *"Rule-of-thumb"* Interaction Variable	-6.67	<0.001	-2.03	0.042

[A] - Variable was not statistically significant in the logistic regression model (p>0.05).

(B) - All participants with > 10 years EA experience had CIH certifications therefore "CIH" was not added to model

(C) - CSP was not analyzed in model since all participants with > 10 years EA with a CSP certification also had a CIH certification

.

Table III.VI and III.VII represent the statistically significant variables ($p<0.05$) from the logistic regression model for analysis of quantitative exposure judgments. Logistic regression models of quantitative judgment accuracy showed positive correlation with greater than 10 years of exposure assessment experience, "High EHS degree", *"DIT Score"*, *"Rule-of-thumb"* data interpretation training, and the number of sample data points available for a judgment as shown in Table III.VI and III.VII. Participants with fewer than 10 years experience indicated negative individual correlations with CIH and CSP certifications and career number of air surveys. Participants with both CIH and CSP certifications ("CIH&CSP") and less than 10 years experience showed positive correlation with quantitative judgment accuracy while greater than 10 years experience showed a negative correlation (Table III.VII). In all quantitative judgment regression models a *"DIT Score"* and *"Rule-of-thumb"* interaction variable showed strong negative correlation (Table III.VI and III.VII).

DISCUSSION

Understanding and identifying the experience, education and training determinants that impact exposure judgment accuracy and bias can be an important step to designing deliberate training practices to raise exposure assessment competence. The participants in this study had a wide range of exposure assessment experience, ranging from less than one year to greater than 30 years. The majority of study participants had at least a Master's degree in a science field (~63%) or a Certified Industrial Hygienist (CIH) certification (~62%) or more than 5 years exposure assessment experience (~63%) as illustrated in Table II and Figures 1 and 2. Binary logistic regression models were constructed to analyze the correlation between the various determinants and exposure judgment accuracy. The education and experience determinants found to positively correlate with both qualitative and quantitative judgment accuracy were *"High EHS Degree"*, *">10 Years EA"* and *"DIT Score"* (Tables III.IV, III.V, III.VI and III.VII). These and other subsequent statistical correlations indicate that several aspects of education, experience and training can significantly impact the accuracy of exposure judgments. In general, these findings related to experience and specific skills agree with studies in other fields showing that properly designed training, experience and education can significantly impact expert performance (Gilovich, 2002, Montgomery, 2005, Ericsson, 2006, Ward, 2007).

Experience

Several experience determinants were collected to look for specific elements that may increase accuracy of exposure judgments. Self-reported experience ranged from general aspects, such as total years of experience documenting exposure assessments and the total number of sampling surveys conducted, to more specific experience determinants of assessing a given task or chemical (Tables III.II and III.III). An attempt was also made to collect a determinant "years since making exposure assessments" to account for hygienists who may have not been actively performing exposure assessments many years and as a result become less accurate. It is possible that this lack of recent practice for participants with many years experience could reduce or obscure actual correlations between years of experience and exposure assessment accuracy. However, a significant group of participants misinterpreted the question, requiring the determinant to be excluded from the models. Even though it appeared that some participants had not actively engaged in exposure assessments for many years, self-reported total years of experience documenting exposure judgments was still found to be significant with qualitative exposure judgments (Tables III.IV and III.VI).

Further analyzing years of experience did produce an interesting finding also seen in other studies. Simon and Chase proposed what is referred to as the "10-Year Rule" based on studies that suggested it takes a minimum of 10 years for a chess player to become an international chess master or grandmaster (Ericsson, 1993, Simon, 1973). This finding was further supported by studies across many fields of expertise, including competitive swimming, music, mathematics, tennis and long distance running. Studies have shown that most fields require a significant accumulation of deliberate practice to achieve a high level of competence regardless of field or domain (Ericsson, 2006). In this study of exposure judgment accuracy, 44% of participants had more than 10 years of experience, allowing for analysis of the 10-Year-Rule on exposure judgment accuracy. When years of experience was analyzed as a binary variable with 10 years as the cut point, more than 10 years experience was significantly correlated for both qualitative and quantitative exposure judgment accuracy (Tables III.IV, III.VI and Figures 3.4 and 3.7). When all qualitative and quantitative exposure judgments made by participants with fewer than 10 years of experience were analyzed across the 4 categories listed in Table III.II, years of experience was not significant for both judgments, offering further support to the 10-Year-Rule finding (Table III.V Analysis #3, Table III.VII Analysis #6).

Specific experience with collecting air samples or performing exposure assessments on a specific agent or task were collected from participants. In most cases these determinants were not found to be specific in any of the models with two exceptions. The number of exposure assessment on a given task or *"TaskNumEAs"* was positively correlated with qualitative judgment accuracy for hygienists with less than 10 years of experience (Table 5 Analysis #3). The variable *"CareerAirSurveys"* was collected to represent

a hygienist's total experience collecting air samples was found to have a negative correlation for the group of participants with less than 10 years of experience in the quantitative regression model (Table III.VII Analysis #7). Intuitively, it may seem that experience with a particular task or chemical would help enhance exposure judgment accuracy. However, recall of this type of specific information for a given task or chemical is likely to be more difficult than estimating the number of years of experience making exposure judgments. Effects from memory and framing have been seen in other studies more formally in a variety of other fields (Sprenger, 2006, Meyer, 2001). Introducing more objective methods for collecting in these specific experience determinants not prone to memory bias would increase likelihood of detecting actual correlations in the binary logistic regression models. More robust methods for defining specific experience than self reporting should be investigated to further study specific experience determinants on exposure judgment accuracy.

Education

Formal post secondary education was reported by participants as *"High EHS Degree"* shown in Table III.II and Figure 3.1. This determinant was meant to capture the highest degree attained in environmental, health and safety (EHS) or science for each participant. Chemistry, biology, biochemistry, engineering and geology were the science degrees reported by participants who did not have a degree that may be directly considered EHS. The *"High EHS Degree"* determinant was significant across all binary logistic regression models for qualitative and quantitative exposure judgments (Tables III.IV- VII). Participants' wide ranges of education levels in this study may have helped illuminate significance of the *"High EHS Degree"* correlation with accuracy in the regression analysis. This finding may seem intuitive; however, a few studies in other fields have indicated that training in mathematics and sciences does not always translate well into domain-specific decision-making processes (Montgomery, 2005). The studies indicated that this type of finding could be due to the complexity and framing of the decision-making process being tested and therefore will not apply uniformly in all cases.

Training

Most would agree that well-designed training increases performance in many domains. It is likely why we see formal certification with training integrated through education and experience in most professional fields, such as medicine, engineering, law and accounting. If a gap within a domain is identified for a profession, specific training and feedback loops can be designed to address the shortcomings of practitioners (Ericsson, 2006). This study focused specifically on exposure judgments, which require a finite but somewhat unknown set of skills to be proficient across many diverse exposure scenarios. The data interpretation rule-of-thumb training in this study focused on a simple heuristic designed to help participants better predict into which AIHA exposure category the 95th percentile fell given a small set of

air samples. This data interpretation rule-of-thumb training has been previously shown to significantly increase accuracy of quantitative exposure judgments but without directly considering other education and experience determinants (Logan, 2009). The data interpretation training was not significant for qualitative exposure judgment accuracy, which seems reasonable as the training only focused on how to interpret data for small data sets independent of any basic characterization information. As mentioned earlier, for training to be effective, it must be focused on the key elements that drive a specific outcome. Qualitative judgments most likely rely on other factors, such as exposure modeling and more specific elements of experience, not captured in this study can be a valuable skill set (Vadali, 2009). Additional studies should be designed to better study predictors of qualitative judgment accuracy including physical chemical modeling and other methods.

A Data Interpretation Test (DIT) was designed to provide feedback to participants and provide a simple measure of data interpretation skills for prospective exposure judgments. A *"DIT Score"* was calculated for each DIT based on the number of correct judgments out of the total judgments on the test. Because the *"DIT Score"* provides some insight into a hygienist's ability to accurately interpret sampling data from a lognormal distribution, it was also used as a determinant in the model prediction for both qualitative and quantitative judgments. The training was highly significant for quantitative exposure judgments and for *"DIT Score"*, which supports earlier findings from Logan et al., using different statistical methods. The *"DIT Score"* is not a perfect measure, but might be analogous to predicting an athlete's performance in real games after studying their performance in practice games. A somewhat surprising finding was that *"DIT Score"* showed positive correlation with qualitative judgment accuracy (Figure 3.5 and Table III.IV) even though these qualitative judgments were made without sampling data. Is it possible that improving *"DIT Score"* will increase qualitative EA accuracy? Industrial hygienists with a higher *"DIT Score"* may have a better understanding of lognormal distributions and upper tail estimations which would lead to more accurate exposure judgments in this context. Additional studies are needed to better understand the relationship between *"DIT Score"* and qualitative judgment accuracy. For all quantitative exposure assessment models, *"DIT Score"* was highly correlated with accurate judgments, having levels of statistical significance above all other determinants in the logistic regression models (Tables III.VI and III.VII). This indicates that *"DIT Score"*, which appears to be a representation of a participant's ability to interpret sampling data, could be used as an important element in training and calibrating occupational hygiene professionals.

When the binary logistic regression analysis was run against quantitative judgments with *"DIT Score"* or "*Rule-of-Thumb*" in the model including all other variables, both *"DIT Score"* and "*Rule-of-Thumb*" were independently found to be significant ($p<0.01$). However, when both were added to the model at the same time, *"DIT Score"* was significant while "*Rule-of-Thumb*" was not significant. A separate analysis showed

that *"DIT Score"* and "*Rule-of-Thumb*" were highly correlated. Because the "*Rule-of-Thumb*" training was designed to influence quantitative exposure judgment skills, it is not a surprise that *"DIT Score"* also increased significantly and, therefore, was highly correlated with "*Rule-of-Thumb*". A solution was implemented to create an interaction variable by multiplying the binary variable "*Rule-of-Thumb*" with the continuous variable *"DIT Score"* to account for effects of the correlation between both variables. [34] When the *"DIT Score"* and *"Rule-of-thumb"* interaction variable was included in the analysis, both *"DIT Score"* and *"Rule-of-Thumb"* became highly significant ($p<0.001$) as a negative correlation shown in Table III.VI and III.VII. This negative correlation with the interaction variable is due to a much stronger correlation between *"DIT Score"* and a participant's judgments before the *"Rule-of-Thumb"* training versus after training. This was verified from an additional logistic regression model showing *"DIT Score"* was highly correlated with accuracy for judgments made before *"Rule-of-Thumb"* training. The same regression model run on judgments made after *"Rule-of-Thumb"* training did not show a correlation between *"DIT Score"* and quantitative judgment accuracy.

Certifications

Professional certifications such as Certified Industrial Hygienist (CIH) and Certified Safety Professional (CSP) require relevant education, experience and acceptable competency as measured by an examination. One may expect that having these certifications may increase exposure judgment accuracy. However, CIH and CSP certifications individually were not found to correlate with qualitative exposure judgment accuracy (Table III.IV and III.V). Each certification individually was negatively correlated with quantitative judgment accuracy for participants with fewer than 10 years experience (Table III.VII Analysis #7). Participants with both certification and more than 10 years experience also showed negative correlation with quantitative judgment accuracy (Table III.VII Analysis #8). However, participants with less than 10 years exposure assessment experience and both CIH and CSP certifications showed a positive correlation with quantitative judgment accuracy (Table III.VII Analysis #7). The negative correlations between professional certifications and quantitative judgment accuracy may suggest that more specific content related to exposure judgments and data interpretation could be added to the exams and study materials. It is important to consider that CIH and CSP cover a broad set of occupational hygiene skills; therefore exposure assessment represents only a very small fraction of competency assessed by each certification.

This study had several limitations which should be addressed in subsequent studies. Participants were self selected in this study and all active occupational hygienists were not made aware or available to participate in this study. This type of selection bias is common to professional judgment elicitation and can sometimes be addressed with adequate funding and improved strategies (Meyer, 2001). The experience and education determinants were ordinal or binary, and subsequent studies should better capture participant information in

continuous variables possibly creating more robust regression models. A final limitation to mention was introduced from selecting the correct exposure category or "truth" using the available 10 to 14 samples for each of the tasks. This assumes the sample data was representative of the true exposure profile for each task and could introduce bias due to limited sample sizes for each of the tasks selected for this study.

The regression modeling methods used assume linear relationships between the education, experience and training determinants and exposure judgment accuracy. This linear model may not accurately represent a hygienist's exposure judgment abilities over time. The relationship between years of experience most likely follows a more polynomial shaped line with accuracy because all "experience" is not the same as it is relative in time rather than absolute in time (Figure 3.10). Not all hygienists are fully engaged in making exposure judgments on a daily basis throughout their career as they change jobs and organizations. A given hygienist could have a very challenging and rich experience depending on the specific job at any point in their career. Considering the idea of "deliberate practice" discussed earlier, a different representation for recent and total exposure judgment experience may provide a better correlation. It is important to remember that many skills may diminish if not used or practiced routinely which is easily seen in athletes where performance is easily measured. Future studies should investigate other methods accounting for current and total exposure judgment experience.

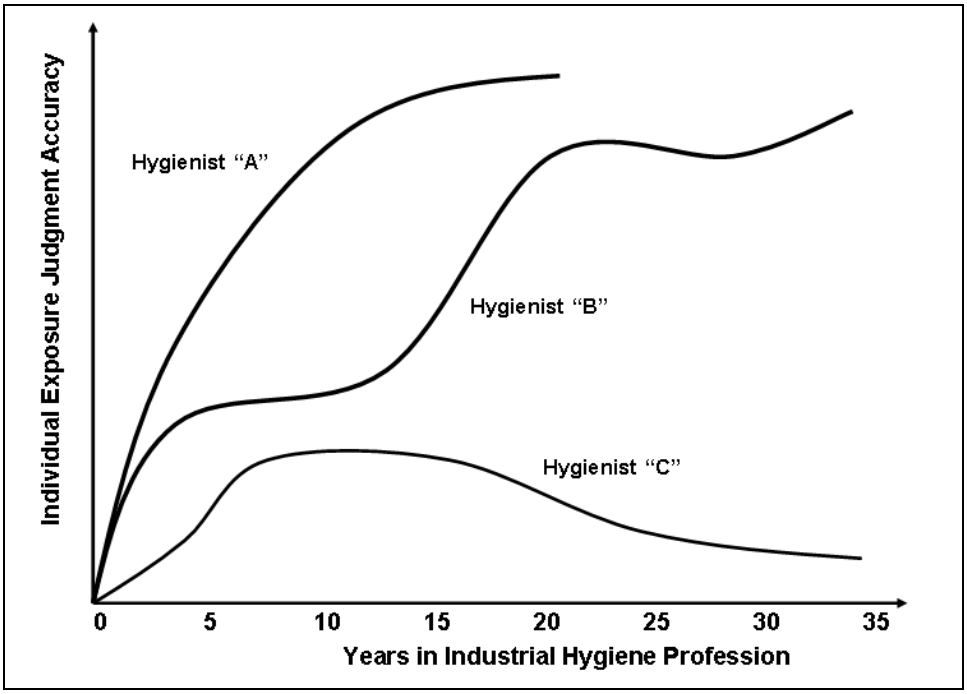

Figure 3.10 - Hypothetical illustration of a non-linear relationship between exposure judgment performance and total years in the industrial hygiene profession for 3 different hygienists

CONCLUSIONS AND RECOMMENDATIONS

This study supports the thesis that formal elements of education, relevant experience and specialized training contribute significantly to exposure judgment accuracy. Implementing well-designed "deliberate practice" mechanisms should support further professional development for occupational and environmental hygienists. Specific training related to exposure assessment techniques such as the data interpretation test and relevant statistical tools can influence data interpretation skills and quantitative exposure judgments. Training on exposure models, chemical behavior, fluid dynamics and physics for occupational hygienists may provide a particular benefit for qualitative exposure assessments when no sampling data is available. The *"DIT Score"* or a similar type of data interpretation metric could become a valuable element in training occupational hygienists analogous to batting practice for professional baseball players. The *"DIT Score"* could be coupled with formal education in science, consistent use of statistical tools and well-designed field experience to significantly enhance professional practice. Since statistical tools are readily available to hygienists, they should always be used when analyzing data, even with small data sets. Efforts should be made in the hygiene profession to continue identifying methods for creating feedback loops and training that further enhances professional practice. As with many other fields, active coaching and mentoring should be considered by all hygienists to strengthen their own skills regardless of education or experience levels and pass on knowledge to the next generation of hygienists. Hygienists who are most interested in enhancing their exposure judgment skills will likely seek every possible opportunity to obtain valuable experience, training, education and mentoring.

This study represents findings from a non-randomly selected group of industrial hygiene professionals, representing approximately 1% of the total profession in the United States. The associations between education, experience and judgment accuracy reported in this study in some ways may apply to the profession as a whole, although they may not apply directly to a specific hygienist. Larger scale studies of this type could better address issues identified in this study and provide for better applicability and more direct interpretation for active occupational hygienists around the world.

Chapter 4 – Performance of Quantitative Exposure Assessment Strategies Including Bayesian Integrated Methods

Quantitative exposure assessment strategies provide the foundation for protecting workers. Properly designed strategies should most often arrive at a correct judgment with the least amount of sampling resources. Simulations were designed to evaluate the performance of several quantitative exposure assessments strategies for different exposure distributions. Bayesian tools are becoming popular and have been included in the simulations for this study along with simple comparison, point estimate and upper confidence limit strategies using minimum sample sizes less than 7 samples. The decision statistic selected for the simulations was the 95th percentile which defines acceptable exceedance fractions by 0.01%, 0.1%, 1%, and unacceptable defined by 10%, 20%, 30% and 50%. Bayesian strategies with using professional judgment were also included to illustrate the impact of an incorrect prior judgment. For acceptable exposure distributions, simple comparison and professional judgment integrated Bayesian strategies showed the highest probability for detecting an acceptable exposure. Bayesian strategies without professional judgment followed by upper confidence limit strategies were least likely to incorrectly define unacceptable exposure distributions as acceptable. Reviewing the different minimum sampling numbers for strategies indicate that Bayesian integrated methods most often arrive a correct decisions with less samples than other strategies. A series of simulations evaluating the Bayesian universe upper geometric standard deviation boundary showed that exposure distributions with high geometric standard deviations reduced that accuracy of Bayesian strategies. Rules of thumb for setting Bayesian universe boundaries are discussed. The results of this study can help design more effective and efficient exposure assessment and management strategies which will hopefully provide a transparent mechanism to strengthen accuracy and bias of exposure judgments.

INTRODUCTION

The AIHA exposure assessment and management strategy is a well known method for comprehensively managing occupational exposure risks by utilizing all available information and data to prioritize and manage exposures in an effective and efficient manner (Figure 4.1). Hygienists categorize workers into similar exposure groups and perform exposure assessments based on the tasks performed, materials, agents presenting potential exposure, and available administrative and engineering exposure controls (Corn and Esman, 1979; Mulhausen and Damiano, 1998; Maidment, 1998; Nicas and Spear, 1992). The strategy recommends systematically collecting all relevant basic characterization information and sampling data needed to arrive at an exposure judgment which determines controls as necessary. This process has been described by Hewett et al as being Bayesian in nature since it incorporates professional judgment and sampling data to arrive at a final judgment. Bayesian statistical methods and tools are now being integrated in many well known exposure assessment strategies (Hewett et al, 2006; Ignacio and Bullock, 2006). An

anticipated promise for Bayesian methods by occupational hygienists is the ability to arrive at correct exposure assessment while utilizing less sampling data therefore allowing more sampling across many other similar exposure groups (SEGs). This broader scope for sampling should ultimately provide a higher level of assurance that all workers are properly protected. Additionally, several of these Bayesian Decision Analysis (BDA) methods allow for incorporation of professional judgment which may be based on experience, exposure models, surrogate or analogous sampling data. Therefore it is important to understand how well Bayesian integrated exposure assessment methods perform for a variety of exposure distributions. This study aims to evaluate the performance of several different exposure assessment strategies including several methods utilizing Bayesian statistics and identify important elements of BDA strategies which lead to better exposure assessment performance (Hewett et al, 2006; Ignacio and Bullock, 2006).

The output of an exposure judgment defines follow up actions related to each exposure assessment. The actions defined are selected based on the most likely exposure control category and the certainty of the assessment (Table IV.I). When sampling data is available, the hygienist would calculate an upper tail decision statistic such as the 95th percentile and select the most appropriate AIHA exposure control category based on the most appropriate occupational exposure limit (OEL). If there is no direct sampling data available for a given SEG (worker-group-agent), the hygienist will utilize available information including their past knowledge and experience to document a qualitative exposure assessment. As qualitative and quantitative exposure assessments are completed, hygienists can prioritize the SEGs based on exposure risk and certainty which define the follow up actions to better understand exposure risks and protect workers (Table IV.I). This iterative and flexible process enables hygienists to focus resources where they are needed most for best understanding and managing exposure risks.

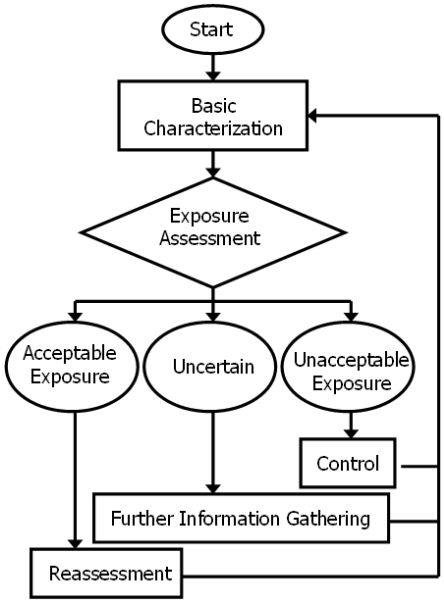

Figure 4.1 – AIHA Exposure Risk Assessment and Management Strategy Flowchart

Table IV.I: AIHA Exposure Category Rating Scheme modified to include exposure Category 0 and follow-up actions. A SEG is assigned an exposure rating by comparing the 95th percentile exposure distribution ($X_{0.95}$) with the full-shift time-weighted average (TWA), Occupational Exposure Limit (OEL) or Short-Term Exposure Limit (STEL) to determine in which category it most likely falls

AIHA Exposure Control Category*	AIHA-Recommended Statistical Interpretation	Control Category General Description	Relative Need for Exposure Monitoring	Possible Follow-up Actions for Exposure Control Category
0	$X_{0.95} \leq 0.01$ OEL	95th percentile of exposures rarely exceeds 1% of the limit.	Very Low	infrequent monitoring to verify qualitative judgment accuracy
1	$X_{0.95} \leq 0.10$ OEL	95th percentile of exposures rarely exceeds 10% of the limit	Low	general hazard communication, infrequent monitoring to verify qualitative judgment accuracy
2	0.10 OEL < $X_{0.95} \leq 0.5$ OEL	95th percentile of exposures rarely exceeds 50% of the limit.	Medium	+ chemical specific hazard communication, periodic exposure monitoring

3	0.5 OEL < $X_{0.95} \leq$ OEL	95th percentile of exposures rarely exceeds the limit.	High	+ medical surveillance, work practice evaluations extensive exposure monitoring
4*	OEL < $X_{0.95}$	95th percentile of exposures exceeds the limit.	Medium	+ respirators, PPE, engineering controls, work practice controls, exposure monitoring to verify respirator protection factors

+ - Include actions from above categories in addition to the current exposure control category

* - Exposure control categories 0-3 are considered "acceptable" while exposure control category 4 is considered "unacceptable"

When samples are collected to support a quantitative exposure assessment, decision rules should be defined to support consistent application of the quantitative exposure assessment strategy. The decision rules should define the minimum number of samples and rules related to statistical confidence which decide when an exposure is considered "acceptable" or "unacceptable". An exposure for an SEG is considered acceptable if the current exposure controls maintain exposures below the threshold of acceptability by having a high likelihood that the decision statistic is below the selected occupational exposure limit (Ignacio and Bullock, 2006). Typically, the 95th percentile ($X_{0.95}$) is used as the decision statistic for most agents however in some select cases hygienists may choose a different upper tail target such as the 90th or 99th percentile (CEN, 1995; HSE, 1999; Mulhausen and Damiano, 1998; Ignacio and Bullock, 2006). With both 'acceptable' and 'unacceptable' exposure judgments, a decision has been made so follow up actions can be implemented which may still include limited verification sampling to ensure good work practices or proper respirator selection. An 'uncertain' quantitative exposure judgment will routinely require more sampling until the judgment threshold is met or a conservative exposure judgment is applied which requires exposure controls to manage the uncertainty. The need for exposure monitoring is different for assessments in each AIHA exposure category. The parameter space area designated as Category 3 ($X_{0.95}$ between 50% and 100% of OEL) is proportionally smaller than all other categories and therefore would have the highest potential for misclassification which could potentially leave workers inadequately protected. The overall risk of misclassification increases as the exposure decision statistic approaches the exposure limit and as a consequence, the number of samples required to arrive at a correct decision increases significantly as the decision statistic gets closer to the exposure limit (Figure 4.2). It is likely that most sampling strategies will be most effective the further the true decision statistic is from the OEL. This same relationship may also be true for professional judgment and may seem intuitive for many hygienists and should be periodically tested (Logan et al, 2009).

Investigators have shown that variability of exposures within a workplace or industry seems to fall into specific ranges depending on the types of operations and technologies utilized in a given industry. For similar agents and operations, most GSDs appear to fall into a range between 2 to 4 (Kromhout et al, 1987). This particular finding is an important element which a Bayesian integrated AIHA strategy can leverage by constraining the range of possible GSDs allowed in the analysis. There are many basic similarities with an AIHA decision strategy only with sampling data using an upper confidence decision statistic and a Bayesian quantitative strategy where professional judgment is not incorporated. A significant difference between the two strategies is that the Bayesian strategy constrains the universe of likely geometric means (GM) and geometric standard deviations (GSD) where as the upper confidence limit methods essentially allow for geometric means and geometric standard deviations that are unbounded. This unbounded condition can generate upper confidence limits that are beyond what is allowed by the laws of physics such as a concentration greater than 1 million parts per million. In contrast, the Bayesian approach can constrain the likely range of geometric means and geometric standard deviations to more reasonable limits. This approach however assumes that the GM and GSD Bayesian universe selected is appropriate however; an error in defining a universe too narrow could result in under estimating exposure risk.

The Bayesian exposure assessment strategies used in this study define decision rules for 'acceptable' in terms of percent probabilities found in the 'unacceptable' AIHA Category 4 (95th percentile greater than 100% of the OEL). Additional simulation inputs include population exceedance fractions (EFs), number of initial samples, acceptable decision thresholds, Bayesian prior and the Bayesian universe boundaries. In particular, the sensitivity of the (1) Bayesian universe GSD upper boundary and (2) an incorrect Bayesian prior on the accuracy of a Bayesian posterior strategy are both important elements to understand when designing strategies and are addressed in this study. The simulations performed in this study were designed for better understanding and design of exposure assessment strategies.

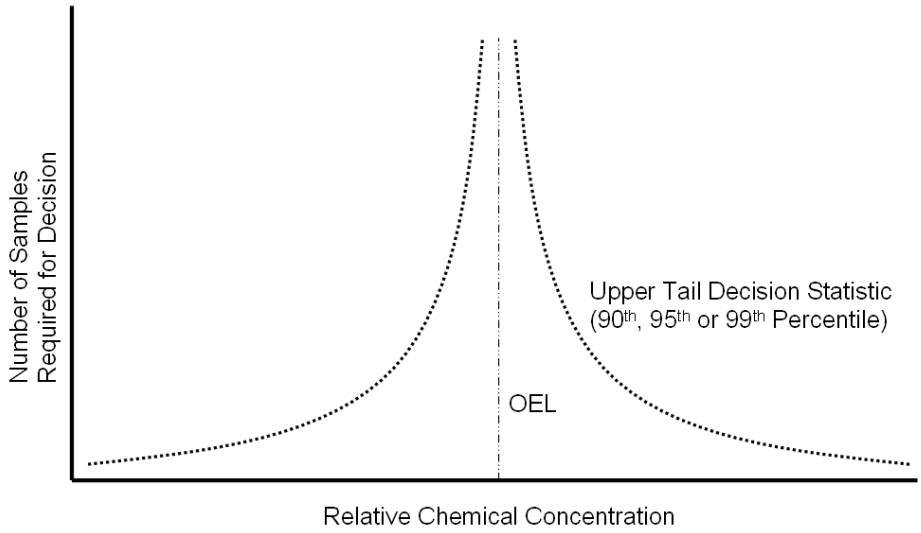

Figure 4.2 – Hypothetical chart illustrating the general relationship between the numbers of samples required for an exposure assessment decision relative to the occupational exposure limit (OEL)

METHODS

In general there were 3 types of sampling strategies evaluated which include setting a minimum sample number for a "Sample Comparison", "Upper Tail Statistic Comparison" and "Bayesian Decision Analysis or BDA". A simple schematic for each type of strategy is illustrated in Figures 4.3, 4.4, and 4.5. The selected strategies were simulated to illustrate how each strategy performed for exceedance fractions (EF) of 0.01%, 0.1%, 1%, 10%, 20%, 30% and 50%. Each EF was run on 3 different exposure populations with different variability for each EF defined by selecting population geometric standard deviations of 2, 3 and 4. For this study the 95th percentile was selected as the decision statistic which defines the line of "acceptable" and "unacceptable". More specifically, it defines the acceptable exposure threshold as no more than 5% exceedance and therefore populations with 0.01%, 0.1% and 1% exceedance fractions were considered "acceptable". Populations with exceedance fractions of 10%, 20%, 30% and 50% were considered "unacceptable" since more than 5% of exposures in the population would exceed the OEL of 100.

Table IV.II describes the strategies for the population exceedance fractions which include the uncertainty rules to arrive at a final decision. Bayesian methods include rules for incorporating uncertainty by selecting a maximum probability in category 4 (>100% of OEL) which were 5% or 10%. The Upper Confidence Limit (UCL) comparison thresholds were set at 90% and 95% which allows for comparison with BDA strategies of 10% and 5% respectively. The Flat Prior BDA Posterior 5% or 10% and Informed Prior BDA Posterior 5% or 10% are illustrated in Figure 4.6 and 4.7 respectively.

Hewett developed a freeware software tool called “EAS Simulator” which allows users to modify various elements of quantitative exposure assessment strategy and show the performance for a range of exposure populations (Hewett 2005). The “EAS Simulator was used to verify the “Compare X0.95 Point Estimate to OEL” and “Compare X0.95 UCL 90% and 95% to OEL” strategy simulations performed in this study using Matlab. The EAS Simulator v2.5.1 has 5 different strategies which allow a user to make various changes to several different strategy parameters including the simulation size, elements of the selected exposure distribution, number of samples, type of sampling campaigns and a number of sampling threshold sensitizing rules (Hewett, 2005). The output for a simulation is a graph where users can easily see the probability that a strategy will classify a given exceedance fraction (EF) as “acceptable” and the average number of samples needed. Users can make changes to the sensitizing rules to design custom strategies which are both effective in properly classifying the exposure and efficient in minimizing the average number of samples needed to make a decision. This tool was used to verify upper confidence limit simulations run in Matlab.

Matlab BDA simulations

Simulation algorithms for each strategy were written in Matlab Student Version 7.8.0.347. The BDA strategies were also written in R to verify the Matlab code output. The “Simple Comparison”, “Compare X0.95 Point Estimate to OEL”, and “Compare X0.95 UCL 95% to OEL” were simulated using the Exposure Assessment Strategies Simulator Version 2.5.1 to verify the Matlab output. Each of the BDA simulations used 4,000 sets of iterations for each of the exceedance fraction and GSD combinations in the study. The Bayesian parameter space for each of the simulations in Tables IV.III-IX were universe geometric mean lower boundary of 0.05, upper boundary of 500, universe geometric standard deviation lower boundary of 1.05 and upper boundary of 4.0. The occupational exposure limit (OEL) for all strategies was 100. Examples of simulation code in Matlab and R can be found in Appendix II.

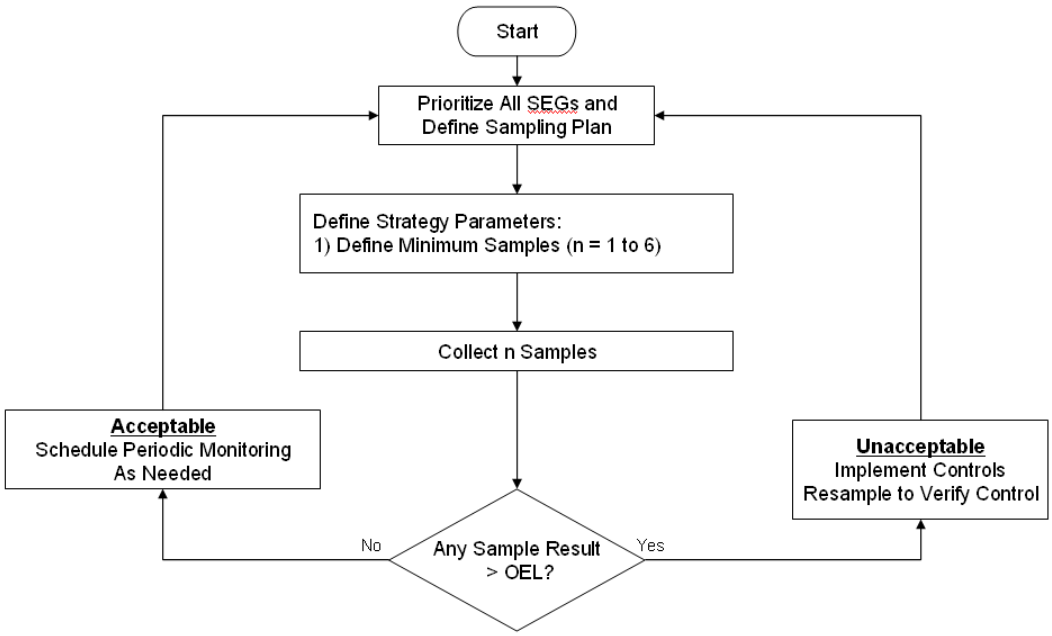

Figure 4.3 - Example Simple Comparison Exposure Assessment Strategy Decision Flowchart

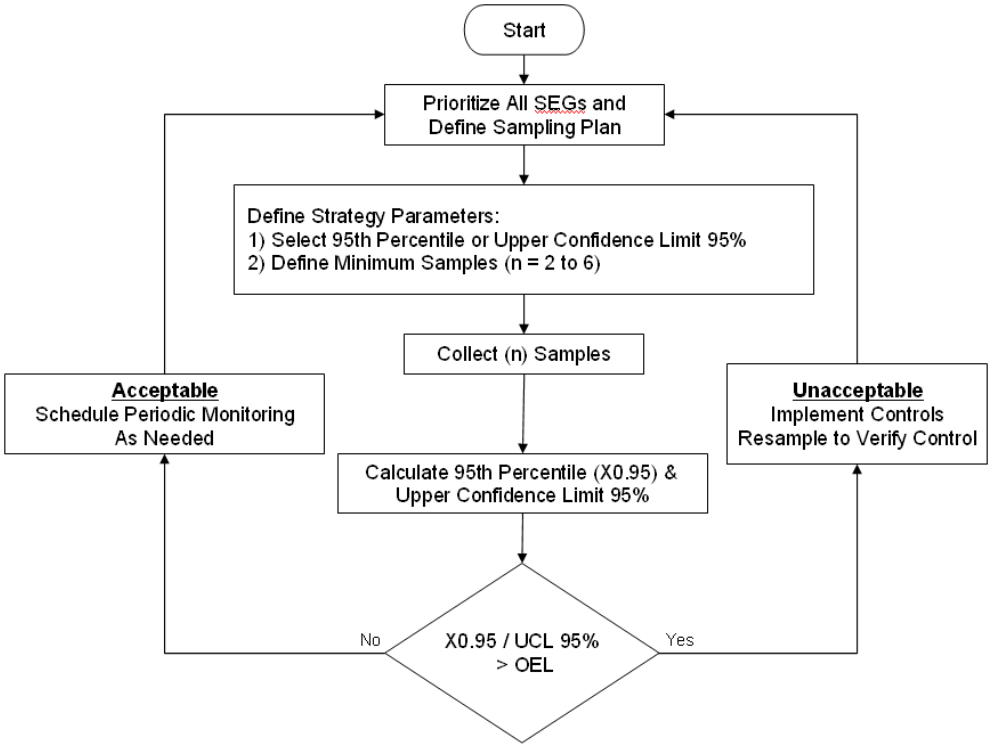

Figure 4.4 - Example Upper Tail Statistic Comparison Exposure Assessment Strategy Decision Flowchart

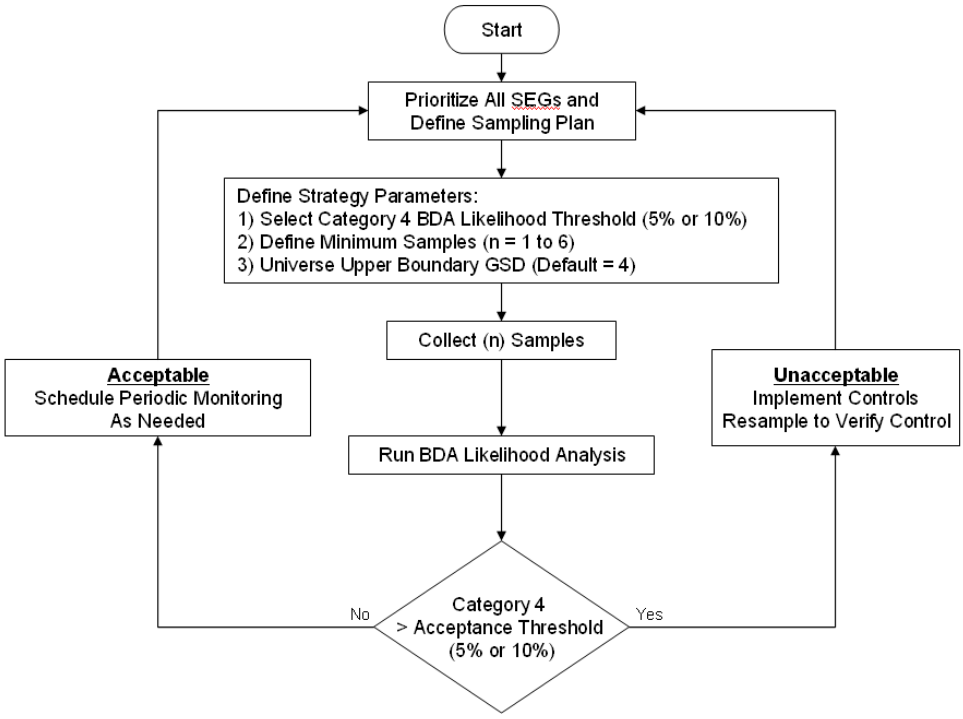

Figure 4.5 - Example Bayesian Quantitative Exposure Assessment Strategy Decision Flowchart

Table IV.II – Description of decision rules used for each strategy simulated

Strategy Name	Description of Strategy for Each Simulation
Simple comparison to 100% of OEL	Define the number of samples (n=1 to 6). Randomly select n samples from the exposure population defined for simulation and compare each sample to the OEL. If any sample is above the OEL, the similar exposure group is considered "unacceptable". If all samples are below the OEL, the SEG is considered "acceptable". Calculate the total percent "acceptable" for all simulations given a specified exposure distribution.
Compare X0.95 point estimate to OEL	Define the number of samples (n=2 to 6). Randomly select n samples from the exposure population defined for simulation and calculate the 95th percentile point estimate [(1)]. If the calculated 95th percentile decision statistic is above the OEL, the similar exposure group is considered "unacceptable". If the calculated 95th percentile decision statistic is below the OEL, the similar exposure group is considered "acceptable". Calculate the total percent "acceptable" for all simulations given a specified exposure distribution.

Compare X0.95 UCL 90% or 95% to OEL	Define the number of samples (n=2 to 6). Randomly select n samples from the exposure population defined for simulation and calculate the 95th percentile point estimate upper confidence .limit [(2)]. If the calculated 95th percentile upper confidence limit (90% or 95%) decision statistic is above the OEL, the similar exposure group is considered "unacceptable". If the calculated 95th percentile upper confidence limit (95%) decision statistic is below the OEL, the similar exposure group is considered "acceptable". Calculate the total percent "acceptable" for all simulations given a specified exposure distribution.
Flat prior BDA posterior 5% or 10%	Define the number of samples (n=1 to 6). Randomly select n samples from the exposure population defined for simulation and calculate the Bayesian Likelihood. If the probability in Category 4 (> OEL) is equal to or greater than 10%, the similar exposure group is considered "unacceptable". If the probability in Category 4 (> OEL) is less than 5% or 10%, the similar exposure group is considered "unacceptable". Calculate the total percent "acceptable" for all simulations given a specified exposure distribution.
Informed prior BDA posterior 5% or 10%	Define the number of samples (n=1 to 6). Randomly select n samples from the exposure population defined for simulation and calculate the Bayesian Posterior using a Professional Judgment Prior initial rating of 2-Well-controlled and High Certainty [(3)]. If the probability in Posterior Category 4 (> OEL) is equal to or greater than 10%, the similar exposure group is considered "unacceptable". If the probability in Posterior Category 4 (> OEL) is less than 10%, the similar exposure group is considered "unacceptable". Calculate the total percent "acceptable" for all simulations given a specified exposure distribution.

[(1)] – The 95th percentile point estimate (X0.95) calculation defined in 3rd Edition requires at least 2 samples

[(2)] – The 95th percentile upper confidence limit (X0.95 UCL 95%) calculation defined in 3rd Edition requires at least 2 samples

[(3)] – Professional Judgment Informed Prior used in simulations was Category 2 (X0.95 between 10% and 50% of OEL) -Well-controlled and High Certainty uses the following probabilities: Category 1 = 20%, Category 2 = 60%, Category 3 = 17%, Category 4 = 3% (Figure 4.8)

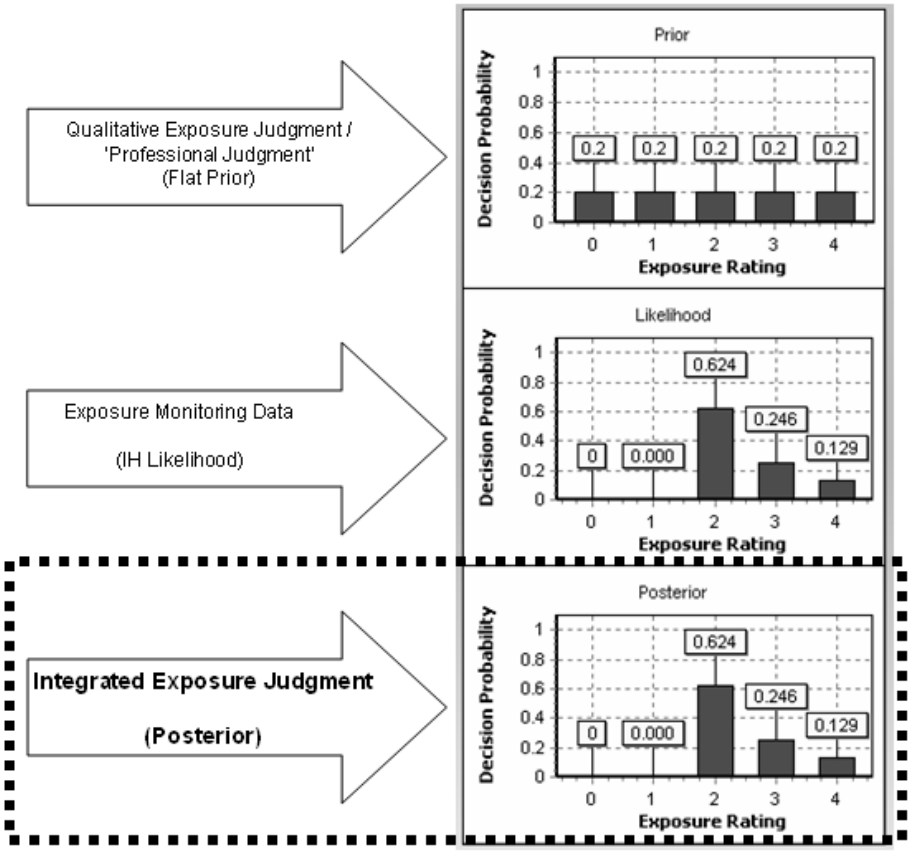

Figure 4.6 – "Bayesian Flat Prior" exposure assessment strategy without professional judgment

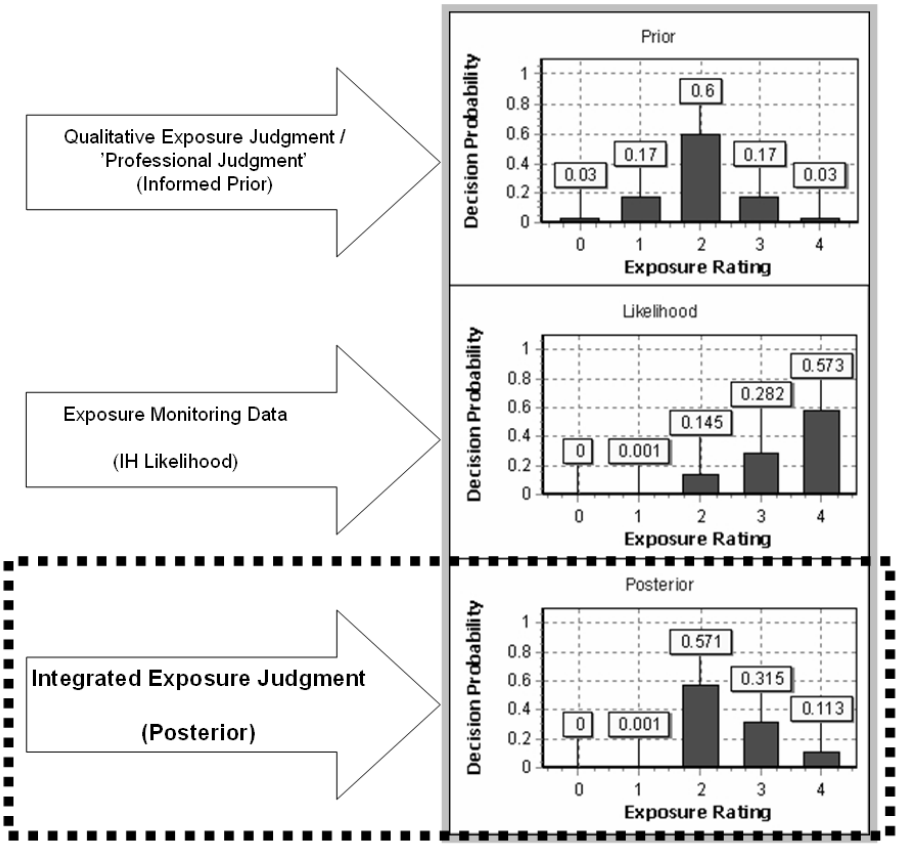

Figure 4.7 – "Bayesian Informed Prior" exposure assessment strategy which incorporates professional judgment

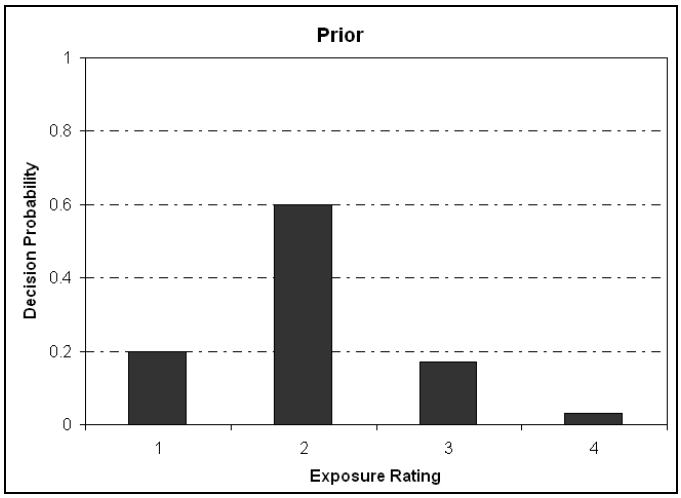

Figure 4.8 – Probabilities used for "Well-controlled and High Certainty" Category 2 Informed Prior in Bayesian Simulations

RESULTS AND DISCUSSION

We don't sample the EFs of 0.01% and 0.1%, that is where professional judgment comes in. Qualitative exposure assessments are often made using professional judgment which is based on available information and data as well has the hygienist's education and experience. Studies have shown that professional judgments made on SEGs without the use of statistical tools and well defined data interpretation rules, routinely produce inaccurate and biased results (Logan, 2009). However, formal elements of education, relevant experience and specialized training did appear to contribute significantly to exposure judgment accuracy and reduction of bias (Chapter 3). The recent availability of Bayesian data analysis tools for exposure data has created new opportunities to better understand exposure judgments and provide transparent feedback mechanisms to increase professional judgment accuracy and reduce bias (Hewett 2006, Logan 2009). In order to properly implement these new Bayesian methods, an understanding of important aspects to implementing a BDA strategy such as managing "parameter space" and defining sample size minimums should be better understood. This study illustrates the performance of several Bayesian integrated sampling strategies compared to other sampling strategies.

Since there have been several Bayesian exposure assessment strategies proposed, the performance of the Bayesian approaches should be simulated and compared to other common strategies (Ramachandran 2001, Ramachancran 2003, Hewett 2006, Logan et al 2009). Hewett et al proposed a Bayesian approach which outputs the probability of an upper tail decision statistic in terms of AIHA exposure categories given a professional judgment prior and sampling data for the likelihood distribution (Hewett 2006). Several investigators have shown the performance of a few common quantitative exposure assessment strategies including a simulated simple comparison to an OEL and upper percentile statistic or upper confidence limit comparisons with an OEL (Tuggle, 1981; Hewett, 2005; Ganser and Hewett, 2010). Hewett has described various simulation methods and tools that can be used to determine how effective a strategy would predict the defined "acceptable" or 'unacceptable' exposure distributions while also illustrating the number of samples generally required to arrive at various decisions. Various input parameters for strategies can be modified to create a highly effective strategy which correctly identifies the true acceptability for all exposure profiles (Figure 4.9). However, this strategy overall would require many more samples than strategies with slightly lower probability for correctly classifying all SEGs (Hewett, 2003, Hewett, 2005). These simulations provide valuable insight to exposure risk managers who are interested in designing strategies which are effective in detecting an unacceptable exposure SEG while efficiently minimizing the number of samples required to make a decision.

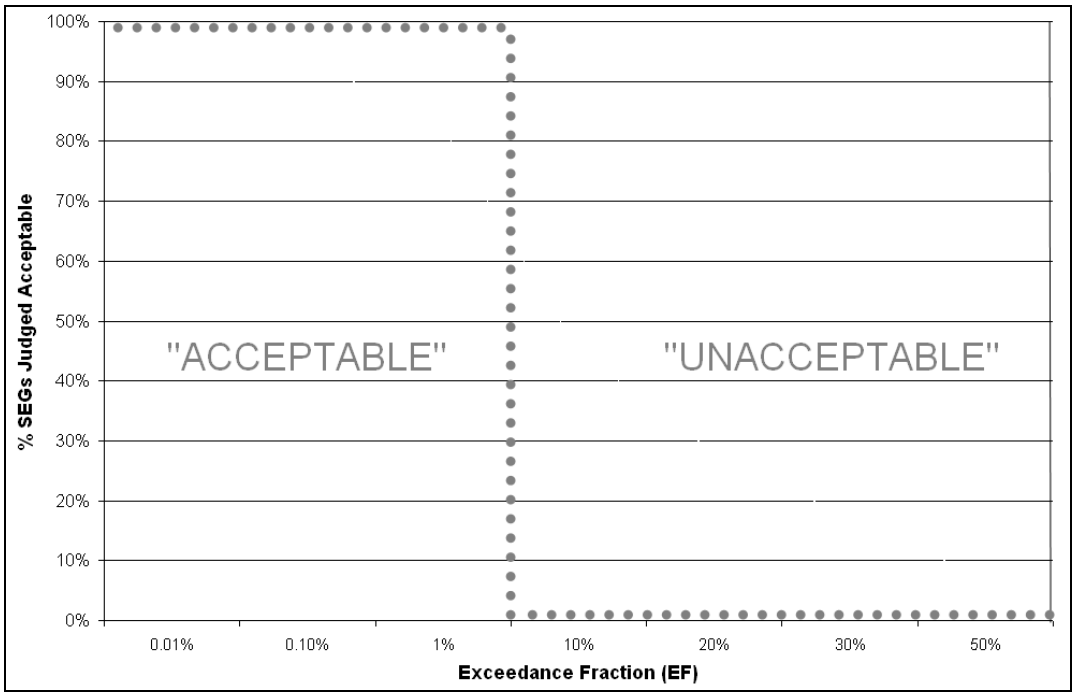

Figure 4.9 – Ideal exposure assessment strategy which correctly identifies acceptability for all exposure profiles using the 95th percentile decision statistic.

Quantitative exposure assessment strategies utilizing Bayesian methods are expected to show promise for arriving at the correct decision with the least number of samples since professional judgment can be specifically incorporated (Hewett, 2006; Logan et al, 2009). The promise of this approach directly depends on the quality of the professional judgment prior used in the Bayesian analysis (Logan, 2009). Constraints can be applied to a Bayesian strategy that can significantly eliminate the impact of professional judgment errors by using an uninformed or "flat" prior and defining the maximum probability in AIHA Category 4 (>100% of OEL) allowed for an acceptable decision (Figure 4.6). This "Flat Prior BDA Posterior" strategy can be a useful method to assess the accuracy of professional judgments which could provide the basis for leveraging a Bayesian Prior (Logan, 2009). An alternative Bayesian Posterior exposure assessment strategy which incorporates the "prior" professional judgment and "likelihood' sampling data into the Posterior decision strategy would likely require less samples to arrive at a final decision (Figure 4.7). For this study a Category 2, high certainty prior was utilized to show the impacts of an incorrect judgment on a Bayesian integrated strategy. In organizations where the accuracy of professional judgments are adequately understood, the Bayesian Posterior exposure assessment strategy could provide a more efficient process while still providing worker protection (Logan, 2009).

When developing an exposure assessment strategy, there are several elements which should be included into the strategy to ensure adequate performance. The strategy should be designed so it can be used to

validate a single exposure judgment while also providing adequate performance over time across all quantitative exposure assessments. Three important elements include the minimum number of samples, GSD universe boundary selections and rules to conclude a SEG is unacceptable. The selection for the minimum sample number is a very important element to ensure that exposure variability of an SEG is properly integrated into the decision making process. At the same time, efficiently arriving at a decision with the least amount of samples is a counter balance to understanding variability.

The results for each simulation are all given as the "Percent Acceptable" for each exposure distribution. For EFs of 0.01%, 0.1% and 1%, this represents the percent of time the strategy would yield a final and correct decision as "acceptable" given the specific rules for each particular strategy (Table IV.II). The residual percent from 100% probability would represent "uncertain" or "unacceptable" which means that either more sampling or exposure controls are required. Strategies with a low percent acceptable indicate a high sampling burden. The "Compare X0.95 UCL 95% to OEL" had a particularly low acceptable percent even for 0.01% EF and 6 samples indicating that many samples would need to be collected to arrive at a final decision. This specific example shows how hygienists would become frustrated trying to implement a strategy where 42% of the time 6 samples would not be enough to conclude an EF of 0.01% was acceptable. For EFs of 10%, 20%, 30% and 50%, the "Percent Acceptable" represents an incorrect or false positive decision. This translates to the percent of SEGs where workers would like not be adequately protected and represents the highest risk of misclassification. When an acceptable SEG is incorrectly rated as "unacceptable", workers would be over protected which mean that additional exposure controls could be implemented when they are not really required. The residual percent from 100% also represents "uncertain" or "unacceptable". The results of these simulations are in the following Tables IV.III – IX and Figures 4.10–15.

Table IV.III – Simulation results for probability determining an SEG as acceptable for an "acceptable" population exceedance fraction of 0.01% with geometric standard deviations of 2, 3 and 4

	Percent Acceptable Decisions (Correct Decision)					
Exceedance Fraction	0.01%					
Population GSD	2					
Population GM	7.59					
Population 95th Percentile	23.7					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	100%	100%	100%	100%	100%	100%
Compare X0.95 Point Estimate to OEL		95%	98%	100%	100%	100%
Compare X0.95 UCL 90% to OEL		21%	39%	55%	68%	79%
Compare X0.95 UCL 95% to OEL		11%	20%	33%	46%	58%
Flat Prior BDA Posterior 10%	47%	67%	77%	85%	90%	93%
Flat Prior BDA Posterior 5%	19%	40%	54%	65%	73%	81%
Cat 2 Informed Prior BDA Posterior 10%	99%	100%	100%	100%	100%	100%
Cat 2 Informed Prior BDA Posterior 5%	98%	98%	99%	99%	100%	100%
	Percent Acceptable Decisions (Correct Decision)					
Exceedance Fraction	0.01%					
Population GSD	3					
Population GM	1.68					
Population 95th Percentile	10.2					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	100%	100%	100%	100%	100%	100%
Compare X0.95 Point Estimate to OEL		95%	98%	100%	100%	100%
Compare X0.95 UCL 90% to OEL		21%	38%	52%	65%	78%
Compare X0.95 UCL 95% to OEL		11%	20%	33%	46%	58%
Flat Prior BDA Posterior 10%	91%	97%	99%	100%	100%	100%
Flat Prior BDA Posterior 5%	79%	93%	97%	99%	100%	100%
Cat 2 Informed Prior BDA Posterior 10%	100%	100%	100%	100%	100%	100%
Cat 2 Informed Prior BDA Posterior 5%	100%	100%	100%	100%	100%	100%
	Percent Acceptable Decisions (Correct Decision)					
Exceedance Fraction	0.01%					

Population GSD	4					
Population GM	0.58					
Population 95th Percentile	5.7					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	100%	100%	100%	100%	100%	100%
Compare X0.95 Point Estimate to OEL		95%	98%	100%	100%	100%
Compare X0.95 UCL 90% to OEL		21%	38%	52%	65%	78%
Compare X0.95 UCL 95% to OEL		11%	20%	33%	46%	58%
Flat Prior BDA Posterior 10%	97%	99%	100%	100%	100%	100%
Flat Prior BDA Posterior 5%	92%	98%	100%	100%	100%	100%
Cat 2 Informed Prior BDA Posterior 10%	100%	100%	100%	100%	100%	100%
Cat 2 Informed Prior BDA Posterior 5%	100%	100%	100%	100%	100%	100%

Table IV.IV – Simulation results for probability determining an SEG as acceptable for an "acceptable" population exceedance fraction of 0.1% with geometric standard deviations of 2, 3 and 4

	Percent Acceptable Decisions (Correct Decision)					
Exceedance Fraction	0.1%					
Population GSD	2					
Population GM	11.7					
Population 95th Percentile	36.6					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	100%	100%	100%	100%	100%	99%
Compare X0.95 Point Estimate to OEL		92%	94%	97%	98%	99%
Compare X0.95 UCL 90% to OEL		19%	29%	40%	49%	58%
Compare X0.95 UCL 95% to OEL		9%	15%	22%	30%	38%
Flat Prior BDA Posterior 10%	25%	40%	49%	54%	62%	62%
Flat Prior BDA Posterior 5%	7%	17%	25%	31%	39%	40%
Cat 2 Informed Prior BDA Posterior 10%	96%	97%	98%	98%	99%	99%
Cat 2 Informed Prior BDA Posterior 5%	91%	91%	93%	94%	95%	96%
	Percent Acceptable Decisions (Correct Decision)					
Exceedance Fraction	0.1%					
Population GSD	3					
Population GM	3.35					

Population 95th Percentile	20.4					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	100%	100%	100%	100%	100%	99%
Compare X0.95 Point Estimate to OEL		92%	94%	97%	98%	99%
Compare X0.95 UCL 90% to OEL		19%	29%	39%	48%	58%
Compare X0.95 UCL 95% to OEL		9%	15%	22%	30%	38%
Flat Prior BDA Posterior 10%	75%	85%	90%	94%	96%	98%
Flat Prior BDA Posterior 5%	57%	72%	80%	86%	91%	94%
Cat 2 Informed Prior BDA Posterior 10%	99%	99%	100%	100%	100%	100%
Cat 2 Informed Prior BDA Posterior 5%	98%	99%	99%	100%	100%	100%
	Percent Acceptable Decisions (Correct Decision)					
Exceedance Fraction	0.1%					
Population GSD	4					
Population GM	1.38					
Population 95th Percentile	13.5					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	100%	100%	100%	100%	100%	99%
Compare X0.95 Point Estimate to OEL		92%	94%	97%	98%	99%
Compare X0.95 UCL 90% to OEL		19%	29%	39%	48%	58%
Compare X0.95 UCL 95% to OEL		9%	15%	22%	30%	38%
Flat Prior BDA Posterior 10%	88%	95%	98%	99%	100%	100%
Flat Prior BDA Posterior 5%	79%	90%	95%	97%	99%	99%
Cat 2 Informed Prior BDA Posterior 10%	99%	100%	100%	100%	100%	100%
Cat 2 Informed Prior BDA Posterior 5%	99%	99%	100%	100%	100%	100%

Table IV.V – Simulation results for probability determining an SEG as acceptable for an "acceptable" population exceedance fraction of 1% with geometric standard deviations of 2, 3 and 4

	Percent Acceptable Decisions (Correct Decision)					
Exceedance Fraction	1%					
Population GSD	2					
Population GM	19.9					
Population 95th Percentile	62					
Number of Samples	1	2	3	4	5	6

Simple Comparison to 100% OEL	99%	98%	97%	96%	95%	94%
Compare X0.95 Point Estimate to OEL		80%	82%	87%	88%	89%
Compare X0.95 UCL 90% to OEL		13%	18%	22%	27%	29%
Compare X0.95 UCL 95% to OEL		7%	9%	12%	14%	17%
Flat Prior BDA Posterior 10%	8%	13%	17%	19%	22%	25%
Flat Prior BDA Posterior 5%	2%	3%	6%	8%	10%	11%
Cat 2 Informed Prior BDA Posterior 10%	86%	84%	80%	79%	80%	78%
Cat 2 Informed Prior BDA Posterior 5%	73%	69%	65%	63%	64%	62%
	Percent Acceptable Decisions (Correct Decision)					
Exceedance Fraction	1%					
Population GSD	3					
Population GM	7.76					
Population 95th Percentile	47					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	99%	98%	97%	96%	95%	94%
Compare X0.95 Point Estimate to OEL		80%	82%	87%	88%	89%
Compare X0.95 UCL 90% to OEL		15%	19%	22%	26%	29%
Compare X0.95 UCL 95% to OEL		7%	9%	12%	14%	17%
Flat Prior BDA Posterior 10%	50%	54%	57%	59%	62%	63%
Flat Prior BDA Posterior 5%	31%	36%	39%	42%	45%	45%
Cat 2 Informed Prior BDA Posterior 10%	94%	94%	96%	95%	95%	96%
Cat 2 Informed Prior BDA Posterior 5%	89%	89%	90%	90%	90%	91%
	Percent Acceptable Decisions (Correct Decision)					
Exceedance Fraction	1%					
Population GSD	4					
Population GM	3.97					
Population 95th Percentile	39					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	99%	98%	97%	96%	95%	94%
Compare X0.95 Point Estimate to OEL		80%	82%	87%	88%	89%
Compare X0.95 UCL 90% to OEL		15%	19%	22%	26%	29%
Compare X0.95 UCL 95% to OEL		7%	9%	12%	14%	17%
Flat Prior BDA Posterior 10%	66%	74%	77%	81%	83%	88%

Flat Prior BDA Posterior 5%	51%	60%	65%	71%	75%	79%
Cat 2 Informed Prior BDA Posterior 10%	95%	97%	98%	98%	98%	99%
Cat 2 Informed Prior BDA Posterior 5%	92%	95%	95%	97%	97%	97%

Table IV.VI – Simulation results for probability determining an SEG as acceptable for an "unacceptable" population exceedance fraction of 10% with geometric standard deviations of 2, 3 and 4

	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	10%					
Population GSD	2					
Population GM	41.1					
Population 95th Percentile	128					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	90%	81%	73%	66%	59%	53%
Compare X0.95 Point Estimate to OEL		53%	45%	40%	37%	34%
Compare X0.95 UCL 90% to OEL		8%	7%	6%	5%	4%
Compare X0.95 UCL 95% to OEL		3%	3%	3%	2%	2%
Flat Prior BDA Posterior 10%	1%	1%	2%	2%	2%	2%
Flat Prior BDA Posterior 5%	0%	0%	0%	1%	1%	1%
Cat 2 Informed Prior BDA Posterior 10%	51%	38%	28%	22%	19%	17%
Cat 2 Informed Prior BDA Posterior 5%	34%	24%	16%	13%	11%	9%
	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	10%					
Population GSD	3					
Population GM	24.5					
Population 95th Percentile	149					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	90%	81%	73%	66%	59%	53%
Compare X0.95 Point Estimate to OEL		53%	45%	40%	37%	34%
Compare X0.95 UCL 90% to OEL		8%	7%	6%	5%	4%
Compare X0.95 UCL 95% to OEL		3%	3%	3%	2%	2%
Flat Prior BDA Posterior 10%	13%	11%	8%	7%	6%	5%
Flat Prior BDA Posterior 5%	5%	4%	3%	3%	2%	2%
Cat 2 Informed Prior BDA Posterior 10%	67%	58%	49%	42%	36%	32%

Cat 2 Informed Prior BDA Posterior 5%	57%	46%	36%	29%	24%	21%
	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	10%					
Population GSD	4					
Population GM	16.9					
Population 95th Percentile	165					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	90%	81%	73%	66%	59%	53%
Compare X0.95 Point Estimate to OEL		53%	45%	40%	37%	34%
Compare X0.95 UCL 90% to OEL		8%	7%	6%	5%	4%
Compare X0.95 UCL 95% to OEL		3%	3%	3%	2%	2%
Flat Prior BDA Posterior 10%	28%	22%	18%	16%	14%	11%
Flat Prior BDA Posterior 5%	16%	13%	10%	9%	7%	6%
Cat 2 Informed Prior BDA Posterior 10%	74%	67%	60%	56%	50%	46%
Cat 2 Informed Prior BDA Posterior 5%	66%	56%	48%	44%	38%	34%

Table IV.VII – Simulation results for probability determining an SEG as acceptable for an "unacceptable" population exceedance fraction of 20% with geometric standard deviations (GSD) of 2, 3 and 4

	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	20%					
Population GSD	2					
Population GM	55.8					
Population 95th Percentile	175					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	80%	64%	51%	41%	33%	26%
Compare X0.95 Point Estimate to OEL		39%	28%	20%	16%	13%
Compare X0.95 UCL 90% to OEL		6%	4%	3%	2%	2%
Compare X0.95 UCL 95% to OEL		3%	2%	2%	1%	1%
Flat Prior BDA Posterior 10%	0%	0%	1%	1%	1%	1%
Flat Prior BDA Posterior 5%	0%	0%	0%	0%	0%	0%
Cat 2 Informed Prior BDA Posterior 10%	35%	20%	13%	8%	6%	5%
Cat 2 Informed Prior BDA Posterior 5%	19%	11%	6%	4%	3%	2%
	Percent Acceptable Decisions					

	(Incorrect Decision)					
Exceedance Fraction	20%					
Population GSD	3					
Population GM	39.7					
Population 95th Percentile	242					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	80%	64%	51%	41%	33%	26%
Compare X0.95 Point Estimate to OEL		39%	28%	20%	16%	13%
Compare X0.95 UCL 90% to OEL		6%	4%	3%	2%	2%
Compare X0.95 UCL 95% to OEL		3%	2%	2%	1%	1%
Flat Prior BDA Posterior 10%	6%	4%	2%	2%	2%	1%
Flat Prior BDA Posterior 5%	2%	1%	1%	1%	1%	0%
Cat 2 Informed Prior BDA Posterior 10%	51%	36%	25%	16%	12%	8%
Cat 2 Informed Prior BDA Posterior 5%	40%	25%	16%	9%	7%	4%
	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	20%					
Population GSD	4					
Population GM	31.1					
Population 95th Percentile	304					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	80%	64%	51%	41%	33%	26%
Compare X0.95 Point Estimate to OEL		39%	28%	20%	16%	13%
Compare X0.95 UCL 90% to OEL		6%	4%	3%	2%	2%
Compare X0.95 UCL 95% to OEL		3%	2%	2%	1%	1%
Flat Prior BDA Posterior 10%	14%	9%	6%	4%	3%	2%
Flat Prior BDA Posterior 5%	8%	5%	3%	2%	1%	1%
Cat 2 Informed Prior BDA Posterior 10%	57%	44%	33%	25%	20%	14%
Cat 2 Informed Prior BDA Posterior 5%	49%	34%	23%	17%	12%	9%

Table IV.VIII – Simulation results for probability determining an SEG as acceptable for an "unacceptable" population exceedance fraction of 30% with geometric standard deviations (GSD) of 2, 3 and 4

	Percent Acceptable Decisions (Incorrect Decision)
Exceedance Fraction	30%

Population GSD	2					
Population GM	69.5					
Population 95th Percentile	217					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	70%	49%	34%	24%	17%	12%
Compare X0.95 Point Estimate to OEL		28%	17%	10%	8%	5%
Compare X0.95 UCL 90% to OEL		4%	2%	2%	1%	0%
Compare X0.95 UCL 95% to OEL		2%	1%	1%	0%	0%
Flat Prior BDA Posterior 10%	0%	0%	0%	0%	0%	0%
Flat Prior BDA Posterior 5%	0%	0%	0%	0%	0%	0%
Cat 2 Informed Prior BDA Posterior 10%	23%	10%	6%	4%	3%	1%
Cat 2 Informed Prior BDA Posterior 5%	12%	5%	3%	2%	1%	1%
	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	30%					
Population GSD	3					
Population GM	56.2					
Population 95th Percentile	343					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	70%	49%	34%	24%	17%	12%
Compare X0.95 Point Estimate to OEL		28%	17%	10%	8%	5%
Compare X0.95 UCL 90% to OEL		4%	2%	2%	1%	0%
Compare X0.95 UCL 95% to OEL		2%	1%	1%	0%	0%
Flat Prior BDA Posterior 10%	4%	1%	1%	1%	0%	0%
Flat Prior BDA Posterior 5%	1%	0%	0%	0%	0%	0%
Cat 2 Informed Prior BDA Posterior 10%	40%	22%	12%	6%	4%	2%
Cat 2 Informed Prior BDA Posterior 5%	29%	14%	7%	4%	2%	1%
	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	30%					
Population GSD	4					
Population GM	48.4					
Population 95th Percentile	473					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	70%	49%	34%	24%	17%	12%

Compare X0.95 Point Estimate to OEL		28%	17%	10%	8%	5%
Compare X0.95 UCL 90% to OEL		4%	2%	2%	1%	0%
Compare X0.95 UCL 95% to OEL		2%	1%	1%	0%	0%
Flat Prior BDA Posterior 10%	8%	4%	2%	1%	1%	0%
Flat Prior BDA Posterior 5%	4%	2%	1%	0%	0%	0%
Cat 2 Informed Prior BDA Posterior 10%	46%	28%	17%	11%	7%	4%
Cat 2 Informed Prior BDA Posterior 5%	37%	20%	11%	7%	4%	2%

Table IV.IX – Simulation results for probability determining an SEG as acceptable for an "unacceptable" population exceedance fraction of 50% with geometric standard deviations (GSD) of 2, 3 and 4

	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	50%					
Population GSD	2					
Population GM	100					
Population 95th Percentile	313					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	50%	25%	12%	6%	3%	2%
Compare X0.95 Point Estimate to OEL		12%	7%	2%	1%	0%
Compare X0.95 UCL 90% to OEL		2%	1%	0%	0%	0%
Compare X0.95 UCL 95% to OEL		1%	0%	0%	0%	0%
Flat Prior BDA Posterior 10%	0%	0%	0%	0%	0%	0%
Flat Prior BDA Posterior 5%	0%	0%	0%	0%	0%	0%
Cat 2 Informed Prior BDA Posterior 10%	9%	2%	1%	1%	0%	0%
Cat 2 Informed Prior BDA Posterior 5%	4%	1%	1%	0%	0%	0%
	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	50%					
Population GSD	3					
Population GM	100					
Population 95th Percentile	609					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	50%	25%	13%	6%	3%	2%
Compare X0.95 Point Estimate to OEL		12%	7%	2%	1%	0%
Compare X0.95 UCL 90% to OEL		2%	1%	0%	0%	0%

Compare X0.95 UCL 95% to OEL		1%	0%	0%	0%	0%
Flat Prior BDA Posterior 10%	1%	0%	0%	0%	0%	0%
Flat Prior BDA Posterior 5%	0%	0%	0%	0%	0%	0%
Cat 2 Informed Prior BDA Posterior 10%	22%	7%	3%	1%	0%	0%
Cat 2 Informed Prior BDA Posterior 5%	14%	4%	1%	1%	0%	0%
	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	50%					
Population GSD	4					
Population GM	100					
Population 95th Percentile	978					
Number of Samples	1	2	3	4	5	6
Simple Comparison to 100% OEL	50%	25%	12%	6%	3%	2%
Compare X0.95 Point Estimate to OEL		12%	7%	2%	1%	0%
Compare X0.95 UCL 90% to OEL		2%	1%	0%	0%	0%
Compare X0.95 UCL 95% to OEL		1%	0%	0%	0%	0%
Flat Prior BDA Posterior 10%	3%	1%	0%	0%	0%	0%
Flat Prior BDA Posterior 5%	1%	0%	0%	0%	0%	0%
Cat 2 Informed Prior BDA Posterior 10%	26%	10%	4%	2%	2%	1%
Cat 2 Informed Prior BDA Posterior 5%	19%	7%	2%	1%	1%	1%

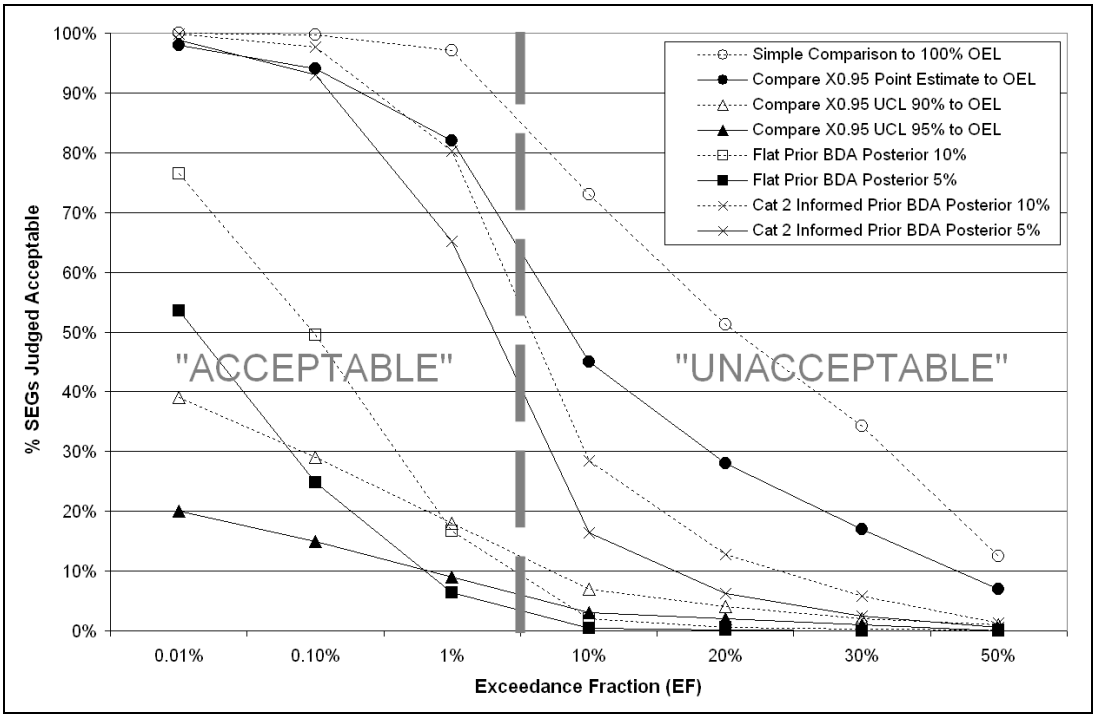

Figure 4.10 - Strategy Simulation Results Using 3 Samples for each Exceedance Fraction with GSD of 2

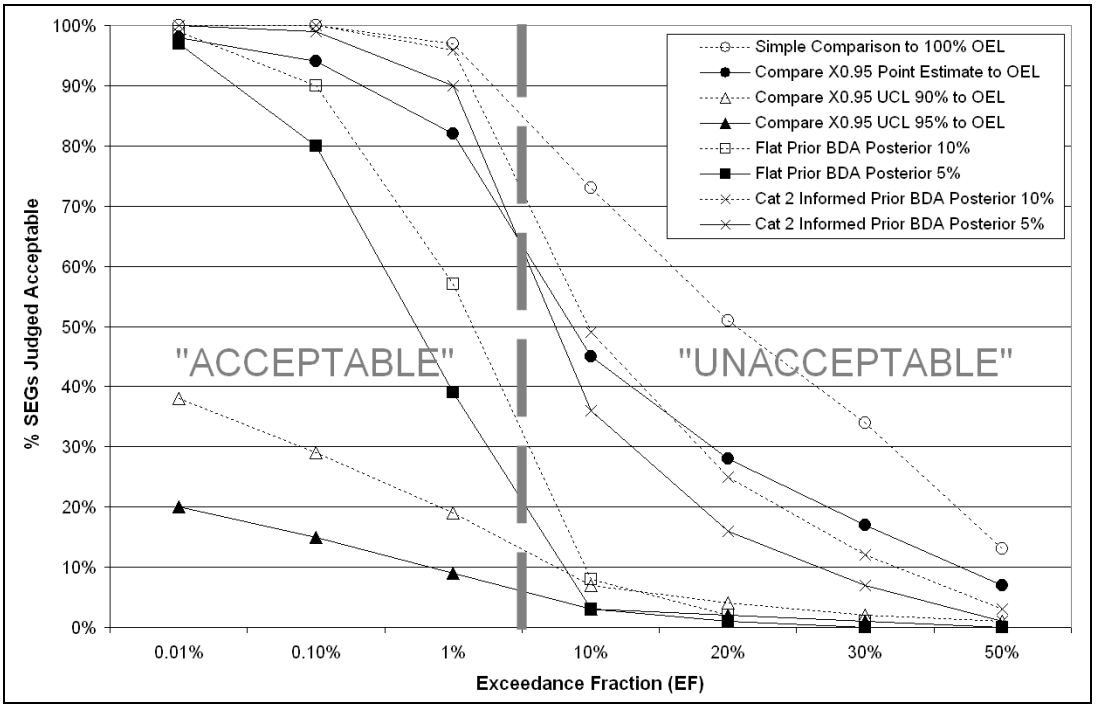

Figure 4.11 - Strategy Simulation Results Using 3 Samples for each Exceedance Fraction with GSD of 3

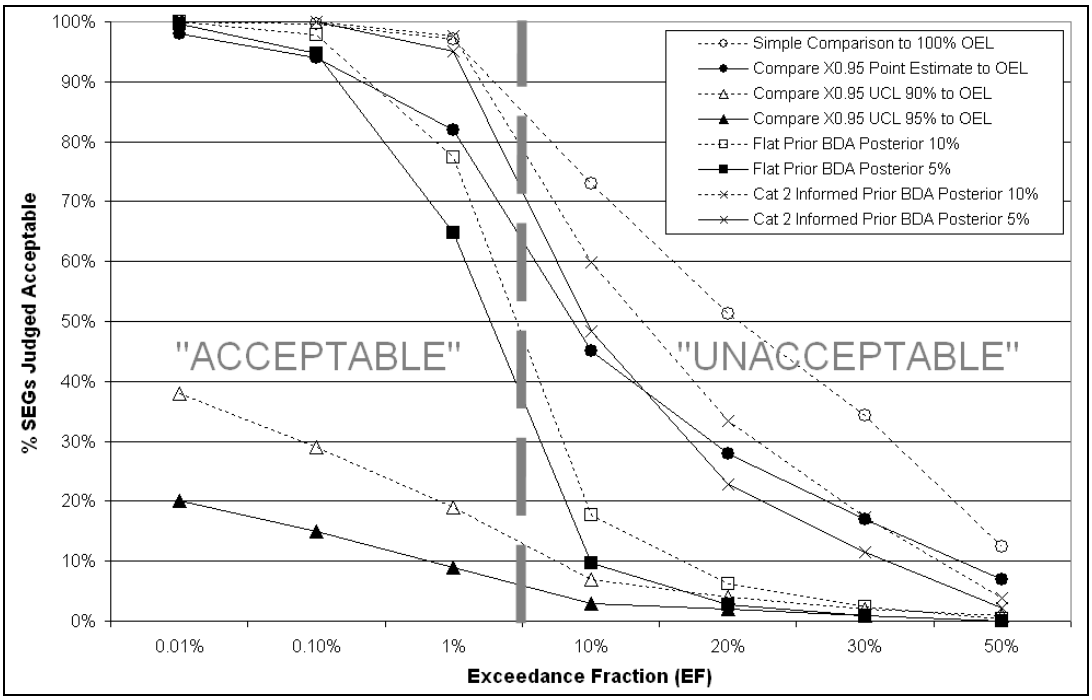

Figure 4.12 - Strategy Simulation Results Using 3 Samples for each Exceedance Fraction with GSD of 4

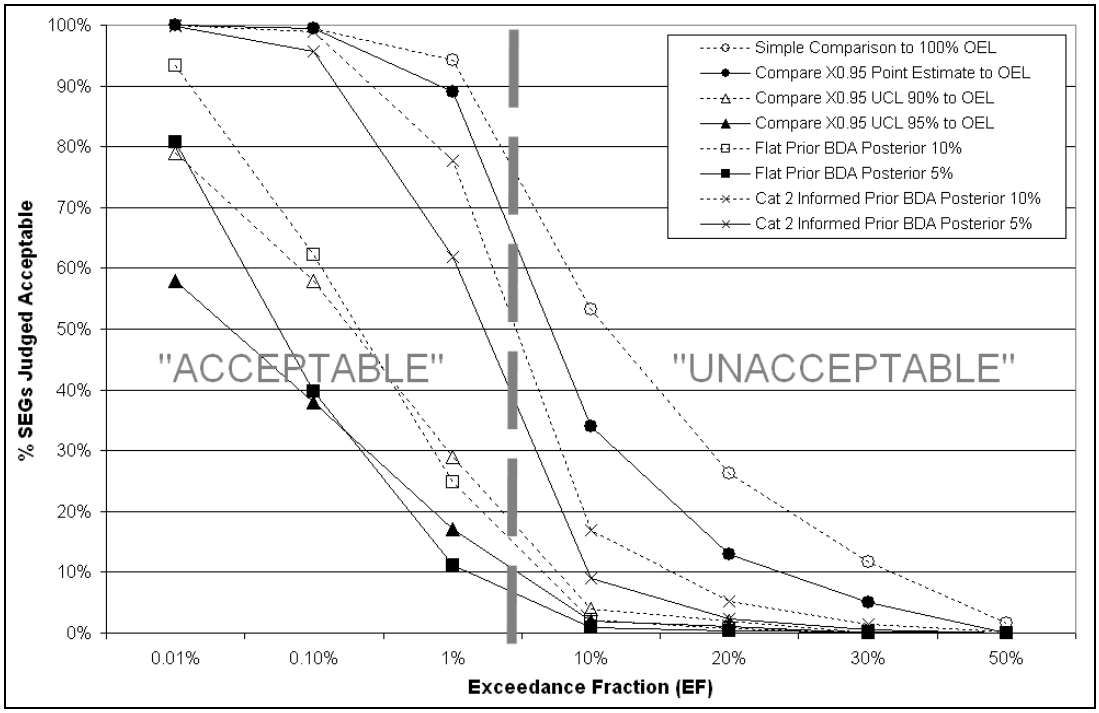

Figure 4.13 - Strategy Simulation Results Using 6 Samples for each Exceedance Fraction with GSD of 2

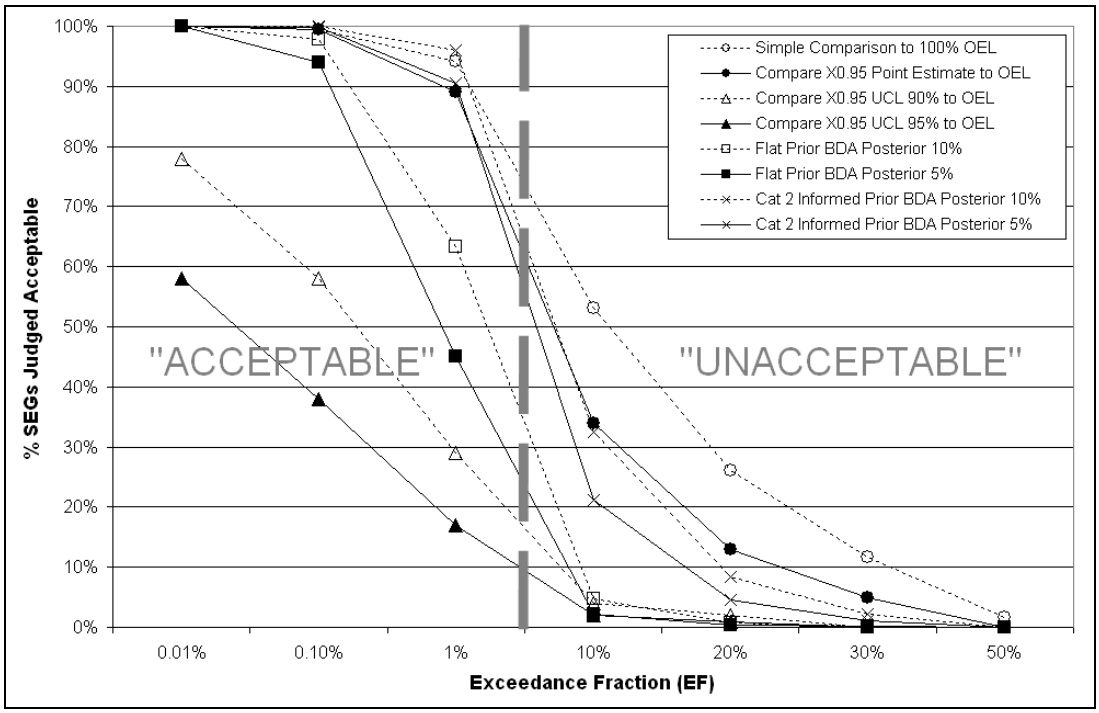

Figure 4.14 - Strategy Simulation Results Using 6 Samples for each Exceedance Fraction with GSD of 3

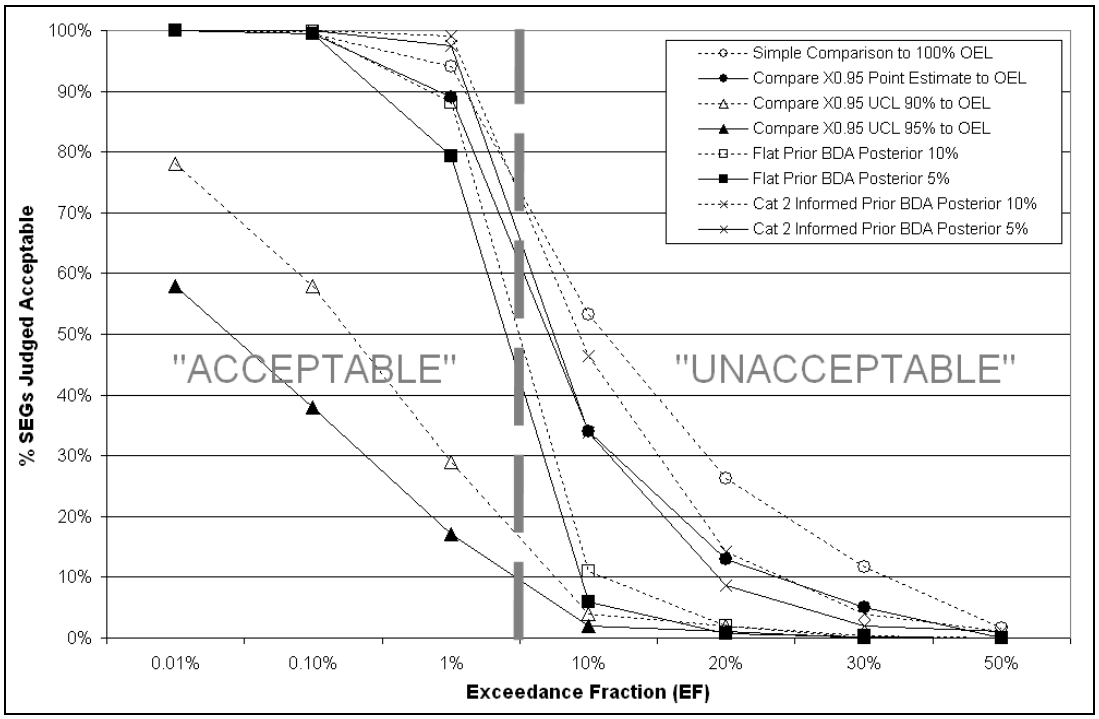

Figure 4.15 - Strategy Simulation Results Using 6 Samples for each Exceedance Fraction with GSD of 4

Simple Comparison to 100% of OEL Strategy

Simple comparisons of sample results to the selected OEL appear easy to apply since no calculations are required. The simple comparison strategy simulated in this study performs well when exposure populations are acceptable and poorly when exposure populations are unacceptable (Tables IV.III-IX, Figures 4.10-15). For acceptable EFs, the simple comparison strategy using 3 samples classifies the SEGs as acceptable at least 97% of the time. The ability to accept an acceptable exposure distribution is better than all other strategies evaluated in this study. However, the strategy has by far the worst performance identifying unacceptable SEGs and therefore should not be used without sensitizing rules or action limits. The simple comparison strategy may also incorrectly infer that a single sample represents the exposure rather than a sample being only one point in a lognormal distribution. This incorrect inference could create a bias which may impact other qualitative and quantitative exposure judgments.

Compare 95th Percentile Point Estimate (X0.95) to OEL Strategy

The Compare X0.95 Point Estimate to OEL appears to perform similar to the simple comparison strategy (Tables IV.III-IX, Figures 4.10-15). It does not perform as well with acceptable exposure profiles but does better than the simple comparison with unacceptable exposure distributions. The strategy using 3 samples for EFs of 10% and 20% the point estimate incorrectly defines the exposure as acceptable 45% and 28% respectively. This strategy does not appear to perform particularly well for any of the simulations when compared to all other strategies.

Compare 95th Percentile Upper Confidence Limits (90% and 95%) Strategies

The upper confidence limit (UCL) strategies performed worse for acceptable exposure profiles than any other strategy tested. This poor performance translates into very high sampling requirements to determine an exposure profile to be acceptable. The UCL strategies however do not have a high rate of incorrectly classifying unacceptable exposure profiles as acceptable. A common critique of upper tail UCL strategies is that the UCL can reach values that are physically impossible such as greater than 1 million parts per million. In addition, UCL strategies require a large number of samples before arriving at a final judgment and seem to require a large number of samples even with EFs at 0.01%.

Bayesian Decision Analysis (BDA) Likelihood 5% and 10% Strategies

BDA methods are becoming much more common in part because of the probability output which lends to user friendly illustration and interpretation. The simulations of the Flat Prior BDA methods show they are able to perform better than most methods for both acceptable and unacceptable exceedance fractions (Tables IV.III-IX, Figures 4.10-15). The BDA method performs much better than the UCL methods for acceptable exposure profiles with the exception of 1% EF and GSD=2. The BDA and UCL methods perform similarly for unacceptable exposure profiles with UCL methods having less false acceptable

percentages with higher GSD exposure profiles. Both places were the UCL method performed better than BDA is a result of the GM and GSD parameter space or universe constraints that are applied in the Bayesian methods. By implementing a process for expanding the BDA universe boundaries, the BDA methods can match and exceed the performance of the UCL methods.

Informed Prior BDA Posterior 5% and 10% Strategies

A well-controlled high certainty category 2 informed prior professional judgment was selected to evaluate how they would perform with both acceptable and unacceptable EFs. For acceptable EFs of 0.01%, 0.1%, and 1%, the informed prior BDA posterior strategy performed much better than the flat prior BDA posterior strategies when population GSDs were lower (Tables IV.III – V, Figures 4.10–15). For unacceptable EFs of 10%, 20%, 30% and 50%, the informed prior BDA strategy performed much worse than the flat prior BDA strategy and almost nearly as bad as the simple comparison strategy (Tables IV.VI–IX, Figures 4.10–15). These simulations suggest that methods should be implemented to determine the general accuracy of qualitative judgments before implementing informed prior BDA strategies.

Initial Judgment Validation Methods

If qualitative judgments and more formal exposure models can be systematically shown to be accurate, the power of Bayesian methods in exposure judgments will be fully realized. A process can be developed to periodically test the accuracy of qualitative judgments made without sampling data or exposure models to support the use of informed priors in a Bayesian exposure assessment strategy. Initial judgment validation methods can be designed using available Bayesian tools along with rules defining minimum number of samples, GSD universe boundary selections and specific rules to conclude a SEG is unacceptable (Hewett et al, 2006; Logan et al, 2009).

Termination Rules

A rule for termination is critical for a hygienist who is focused on applying an exposure assessment strategy in a workplace. Hygienists need both rules for judging SEGs as “acceptable” or “unacceptable” from “uncertain”. Specific rules should be defined up front in each strategy so that a final decision of “unacceptable” is selected and efforts are directed toward implementing exposure controls. Termination rules are an important part of quickly focusing efforts on controls rather than additional sampling. For a simple comparison strategy, a single sample above the OEL or other threshold such as 50% of the OEL could be incorporated into a strategy to increase efficiency. For upper percentile comparison strategies, it is more difficult to create and implement termination strategies since upper confidence limits are non-linear. Selecting a termination threshold would likely not seem intuitive and likely create implementation challenges (Figure 4.16). Termination strategies for Bayesian strategies appear to be more intuitive than upper confidence limit strategies since probabilities would be used as thresholds. Using a flat prior

Bayesian strategy, a hygienist could termination threshold such as “greater than 50% in category 4” which is very easy to communicate (Figure 4.16). An example communication could be, “there is a greater than 50% chance that this SEG is unacceptable” allowing the hygienist to focus efforts for controls. An illustration of this modified strategy is depicted in Figure 4.17.

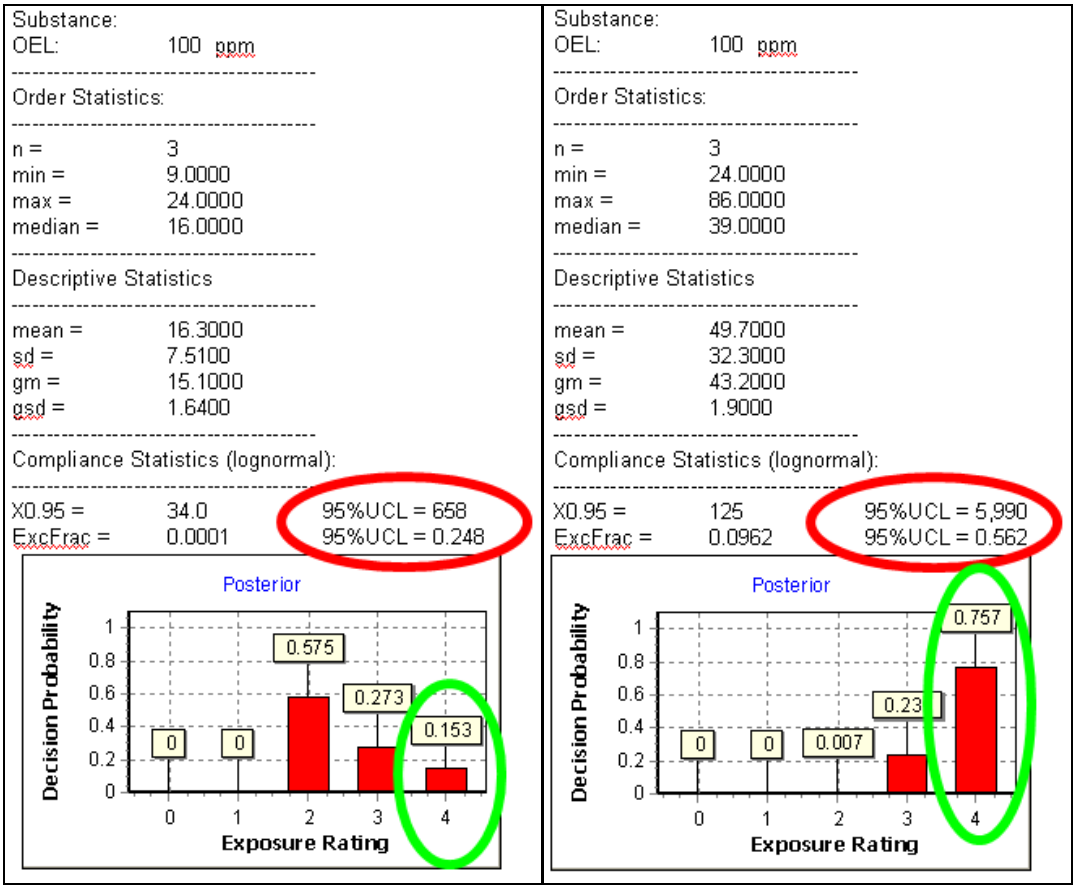

Figure 4.16 – Two example scenarios to illustrate comparison of BDA and UCL termination rules

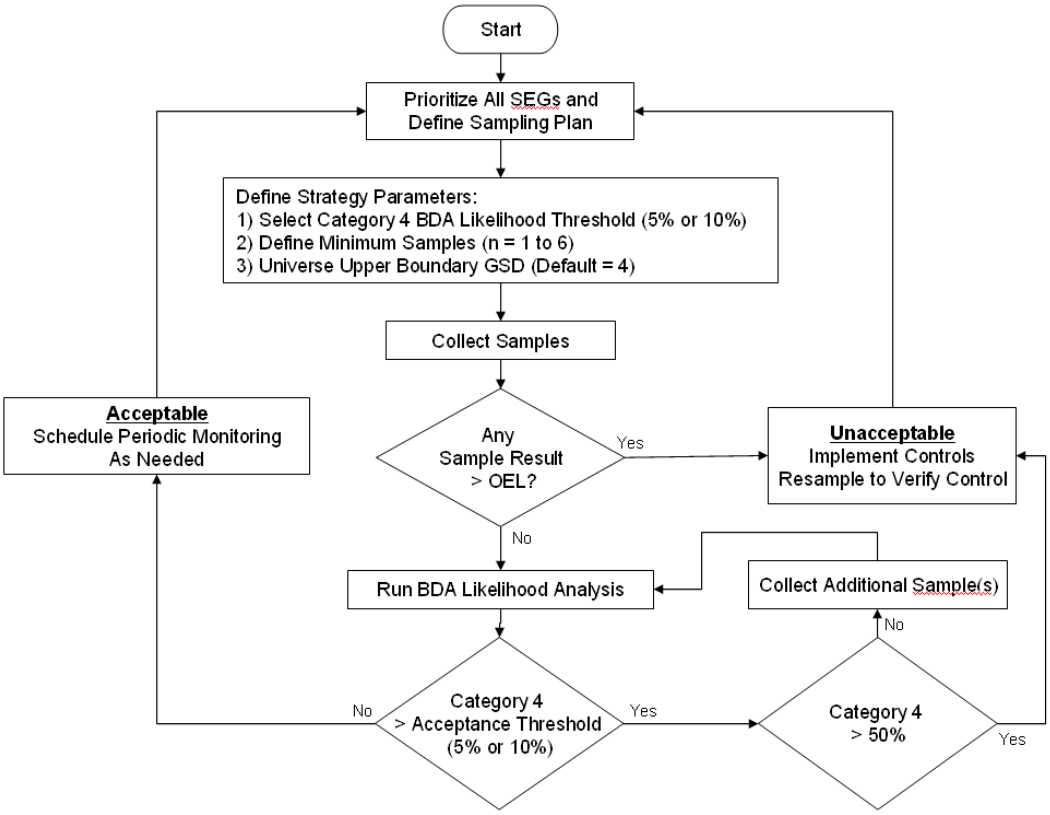

Figure 4.17 - Modified Bayesian Quantitative Exposure Assessment Strategy Decision Flowchart

A series of simulations were run on flat prior BDA posterior strategies with 3 samples across all exceedance fractions with a population GSD of 3 in order to see how much of the residual "uncertain" judgments would have a final decision of "unacceptable" using a 50% termination threshold. The simulations show that for unacceptable exceedance fractions, a significant proportion would have category 4 probability above the 50% termination threshold (Table IV.X). For an EF of 10%, around 50% of the time the SEG would be deemed "unacceptable" with only 3 samples. The majority of the remaining would be deemed "uncertain" with 3% or 8% of the balance being "acceptable" for the 5% and 10% threshold strategies respectively.

Table IV.X – Percent of simulations greater than 50% probability in Category 4 using a Flat Prior BDA Posterior Strategy with 3 samples and a GSD = 3 across all exceedance fractions

Exceedance Fraction	Flat Prior BDA Posterior 3 Sample Strategy % of SEGs with Category 4 >50%
0.01%	0%
0.1%	0%
1%	6%
10%	50%
20%	75%
30%	88%
50%	97%

The BDA strategy performance decreases as the exposure population GSD becomes close to the BDA GSD upper GSD boundary. Strategies should include rules which modify BDA universe GSD upper boundary to ensure good performance. Simulations in Tables IV.XI-XVII illustrate how the performance of the BDA Flat Prior methods is enhanced by increasing the GSD upper parameter space boundary. For 10% EF and 3 samples, when the GSD upper universe boundary is expanded from 4 to 6, the performance comes in line with the UCL methods (Tables IV.V -XIV). Expanding the GSD upper boundary to 10 increases the performance beyond the UCL method for that particular exposure distribution. In general, adjusting the BDA GSD upper boundary to be 2 larger than the sample GSD appears to provide reasonable assurance that variability is accounted for in each analysis.

Table IV.XI – Bayesian Upper GSD Universe Boundary Simulation results for probability determining an SEG as acceptable for an "acceptable" population exceedance fraction of 0.01% and GSD of 4

	Percent Acceptable Decisions (Correct Decision)					
Exceedance Fraction	0.01%					
Population GSD	4					
Population GM	0.58					
Population 95th Percentile	5.7					
BDA GSD Universe Upper Boundary	4					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	97%	99%	100%	100%	100%	100%
Flat Prior BDA Posterior 5%	92%	98%	100%	100%	100%	100%

BDA GSD Universe Upper Boundary	6					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	90%	97%	99%	100%	100%	100%
Flat Prior BDA Posterior 5%	78%	91%	96%	99%	99%	100%
BDA GSD Universe Upper Boundary	10					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	73%	85%	92%	95%	98%	99%
Flat Prior BDA Posterior 5%	45%	67%	80%	87%	93%	95%

Table IV.XII – Bayesian Upper GSD Universe Boundary Simulation results for probability determining an SEG as acceptable for an "acceptable" population exceedance fraction of 0.1% and GSD of 4

	Percent Acceptable Decisions (Correct Decision)					
Exceedance Fraction	0.10%					
Population GSD	4					
Population GM	1.38					
Population 95th Percentile	13.5					
BDA GSD Universe Upper Boundary	4					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	88%	95%	98%	99%	100%	100%
Flat Prior BDA Posterior 5%	79%	90%	95%	97%	99%	99%
BDA GSD Universe Upper Boundary	6					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	74%	85%	89%	93%	96%	97%
Flat Prior BDA Posterior 5%	55%	71%	79%	86%	90%	93%
BDA GSD Universe Upper Boundary	10					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	48%	63%	69%	75%	80%	85%
Flat Prior BDA Posterior 5%	22%	40%	47%	55%	61%	67%

Table IV.XIII – Bayesian Upper GSD Universe Boundary Simulation results for probability determining an SEG as acceptable for an "acceptable" population exceedance fraction of 1% and GSD of 4

	Percent Acceptable Decisions (Correct Decision)
Exceedance Fraction	1%

Population GSD	4					
Population GM	3.97					
Population 95th Percentile	39					
BDA GSD Universe Upper Boundary	4					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	66%	74%	77%	81%	83%	88%
Flat Prior BDA Posterior 5%	51%	60%	65%	71%	75%	79%
BDA GSD Universe Upper Boundary	6					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	44%	51%	54%	55%	57%	64%
Flat Prior BDA Posterior 5%	25%	32%	36%	37%	39%	46%
BDA GSD Universe Upper Boundary	10					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	20%	28%	29%	29%	31%	33%
Flat Prior BDA Posterior 5%	7%	11%	13%	14%	16%	17%

Table IV.XIV – Bayesian Upper GSD Universe Boundary Simulation results for probability determining an SEG as acceptable for an "unacceptable" population exceedance fraction of 10% and GSD of 4

	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	10%					
Population GSD	4					
Population GM	16.9					
Population 95th Percentile	165					
BDA GSD Universe Upper Boundary	4					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	28%	22%	18%	16%	14%	11%
Flat Prior BDA Posterior 5%	16%	13%	10%	9%	7%	6%
BDA GSD Universe Upper Boundary	6					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	13%	9%	7%	6%	6%	4%
Flat Prior BDA Posterior 5%	5%	4%	3%	3%	3%	2%
BDA GSD Universe Upper Boundary	10					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	3%	3%	3%	3%	3%	3%

Flat Prior BDA Posterior 5%	1%	0%	1%	1%	1%	1%

Table IV.XV – Bayesian Upper GSD Universe Boundary Simulation results for probability determining an SEG as acceptable for an "unacceptable" population exceedance fraction of 20% and GSD of 4

	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	20%					
Population GSD	4					
Population GM	31.1					
Population 95th Percentile	304					
BDA GSD Universe Upper Boundary	4					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	14%	9%	6%	4%	3%	2%
Flat Prior BDA Posterior 5%	8%	5%	3%	2%	1%	1%
BDA GSD Universe Upper Boundary	6					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	6%	3%	2%	2%	1%	1%
Flat Prior BDA Posterior 5%	2%	1%	1%	1%	0%	0%
BDA GSD Universe Upper Boundary	10					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	1%	1%	1%	1%	1%	0%
Flat Prior BDA Posterior 5%	0%	0%	0%	0%	0%	0%

Table IV.XVI – Bayesian Upper GSD Universe Boundary Simulation results for probability determining an SEG as acceptable for an "unacceptable" population exceedance fraction of 30% and GSD of 4

	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	30%					
Population GSD	4					
Population GM	48.4					
Population 95th Percentile	473					
BDA GSD Universe Upper Boundary	4					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	8%	4%	2%	1%	1%	0%
Flat Prior BDA Posterior 5%	4%	2%	1%	0%	0%	0%

BDA GSD Universe Upper Boundary	6					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	3%	1%	1%	1%	0%	0%
Flat Prior BDA Posterior 5%	1%	0%	0%	0%	0%	0%
BDA GSD Universe Upper Boundary	10					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	1%	0%	0%	1%	0%	0%
Flat Prior BDA Posterior 5%	0%	0%	0%	0%	0%	0%

Table IV.XVII – Bayesian Upper GSD Universe Boundary Simulation results for probability determining an SEG as acceptable for an "unacceptable" population exceedance fraction of 50% and GSD of 4

	Percent Acceptable Decisions (Incorrect Decision)					
Exceedance Fraction	50%					
Population GSD	4					
Population GM	100					
Population 95th Percentile	978					
BDA GSD Universe Upper Boundary	4					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	3%	1%	0%	0%	0%	0%
Flat Prior BDA Posterior 5%	1%	0%	0%	0%	0%	0%
BDA GSD Universe Upper Boundary	6					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	1%	0%	0%	0%	0%	0%
Flat Prior BDA Posterior 5%	0%	0%	0%	0%	0%	0%
BDA GSD Universe Upper Boundary	10					
Number of Samples	1	2	3	4	5	6
Flat Prior BDA Posterior 10%	0%	0%	0%	0%	0%	0%
Flat Prior BDA Posterior 5%	0%	0%	0%	0%	0%	0%

CONCLUSIONS AND RECOMMENDATIONS

The simulations in this study were designed to help exposure risk managers understand the performance of various strategies when modifying input parameters. The simulations on simple strategies illustrate how poorly they detect an unacceptable exposure as compared with the upper tail and Bayesian strategies. When considering sampling efficiency and correct exposure judgment effectiveness, the "Flat Prior BDA Posterior" performed better overall than nearly all of the strategies for the EFs used in these simulations. If one would want to ensure less than 10% incorrect classification rates for unacceptable exposure profiles, a flat prior BDA strategy incorporating a minimum sample number of at least 3 along with rules for expanding the BDA GSD upper parameter space boundaries and setting a termination rule of 50% would be adequate. Organizations should periodically review SEGs to determine the actual distribution of sample GSDs for their operations thereby giving better insight to selecting BDA GSD universe boundaries.

The performance of Bayesian methods is of particular interest since these methods are now becoming informally and formally incorporated into exposure assessment strategies. These simulations illustrate how Bayesian integrated methods can give better overall performance than other common strategies if rules for minimum sample number, acceptable and unacceptable thresholds are properly defined and followed. Exposure assessment professionals should use available tools to test and design exposure assessment strategies that allow for validation of professional judgments to ensure better understanding of overall exposure judgment performance.

Chapter 5 – Conclusions and Future Direction

Occupational hygienists perform countless exposure judgments throughout their career which ultimately determine how well people are protected from adverse health effects. It appears that hygienists are like other professionals and prone to errors in exposure judgments (Logan et al, 2009). As seen in other professions, these judgments are likely to also be made using a complex blend of cognitive rules and beliefs combined with emotional and unconscious mechanisms or intuition (Zsambok and Klein, 1997; Gigerenzer and Todd, 1999; Kahneman and Tversky, 2000; Gilovich et al, 2002; Schneider and Shanteau, 3003). The complex mechanisms are most certainly used when making exposure judgments and appear to align well with Bayesian integrated exposure assessment strategies. It appears that exposure judgments are not only inaccurate but also can be biased which in some cases could put workers at higher risk. Fortunately, Bayesian integrated strategies offer a transparent mechanism for direct feedback and systematic processes to evaluate many aspects of exposure judgments. In addition, it appears that focused training and other elements of deliberate practice could provide additional mechanisms for increasing accuracy and eliminating bias in exposure judgments. The central elements of this research point to both challenges and opportunities for exposure assessment professionals.

- Exposure judgments are more accurate than random chance but can be biased such that some workers could be under protected.
- Data interpretation training can increase accuracy in both qualitative and quantitative exposure judgments while reducing bias in quantitative exposure judgments.
- Increased levels of education and experience correlate with exposure judgment accuracy however professional certifications do not appear to significantly impact accuracy.
- Bayesian quantitative exposure assessment strategies utilizing uninformative or flat prior judgments offer an incremental increase in overall efficiency and effectiveness compared to other common strategies.

The feedback mechanisms by using Bayesian integrated methods will require metrics to help exposure assessment professionals identify when a particular judgment does not adequately match the available data for a given SEG. In this study, several metrics were considered when testing exposure judgment accuracy (Equations 5.1, 5.2, 5.3). However the percent category agreement (PCA) metric was selected because category selection better correlates with follow up actions and the PCA metric was significantly easier to apply to the data collected. As seen in Equation 5.2, the PCA metric only considers the highest category (HC) rather than using probabilities across each of the categories (CatT). However, in order to create a metric that can be used for feedback, it needs to incorporate elements of correct category and the amount of probability overlap in each category.

- *Fractional Overlap (FO) -* $$FO = \left(1 - \sum_{i=1}^{4} \left(CatT_i - CatJ_i\right)^2\right)$$ (Equation 5.1)

- *Percent Category Agreement (PCA) -* $$PCA = 1 - \left(\frac{\left|(HC_T - HC_J)\right|}{3}\right)$$ (Equation 5.2)

- *Probability Overlap (PO) -* $$PO = 1 - \sum_{i=1}^{4} \left|CatT_i - CatJ_i\right|$$ (Equation 5.3)

Exposure judgments with extremely high or low agreement are obvious to discern from probability charts however as judgments begin to match, the task becomes more difficult (Figures 5.1 and 5.3). A significant number of exposure judgments were constructed in an iterative fashion for very high agreement all the way down to low agreement so that each metric could be evaluated (Figures 5.2). Combinations of metrics were evaluated to attempt to create a metric which incorporated both category agreement and probability overlap.

IH Bayesian Prior Likelihood Agreement Metric (BPLA) - To compare prior judgments with likelihood charts, one could compute a probability overlap and category agreement utilizing the PO and PCA equations above. If the exposure judgments between a participant and "Truth" are identical, then the probability and category overlap is unity.

$$BPLA = 0.3 * PO + 0.7 * PCA$$ (Equation 5.4)

The "Bayesian Prior Likelihood Agreement" or BPLA metric is proposed to help better understand the amount of agreement between the IH Prior and IH Likelihood to clearly identify when a large discrepancy exists where follow up is needed. This equation created for the BPLA metric is a combination of PCA and PO equations above. A perfect match for the selected exposure category and probabilities assigned in each category produces a BPLA metric equal to 1.0 (Figure 5.3a). A perfect mismatch in both category and probability will produce the lowest possible BPLA metric (Figure 5.3b).

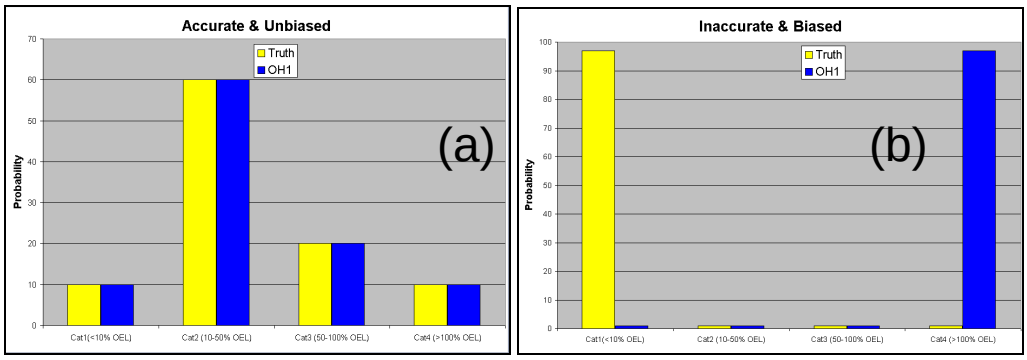

Figure 5.1. Charts of exposure judgments (OH1) versus "Truth" illustrating a perfectly accurate judgment (a) and a totally inaccurate and biased judgment (b).

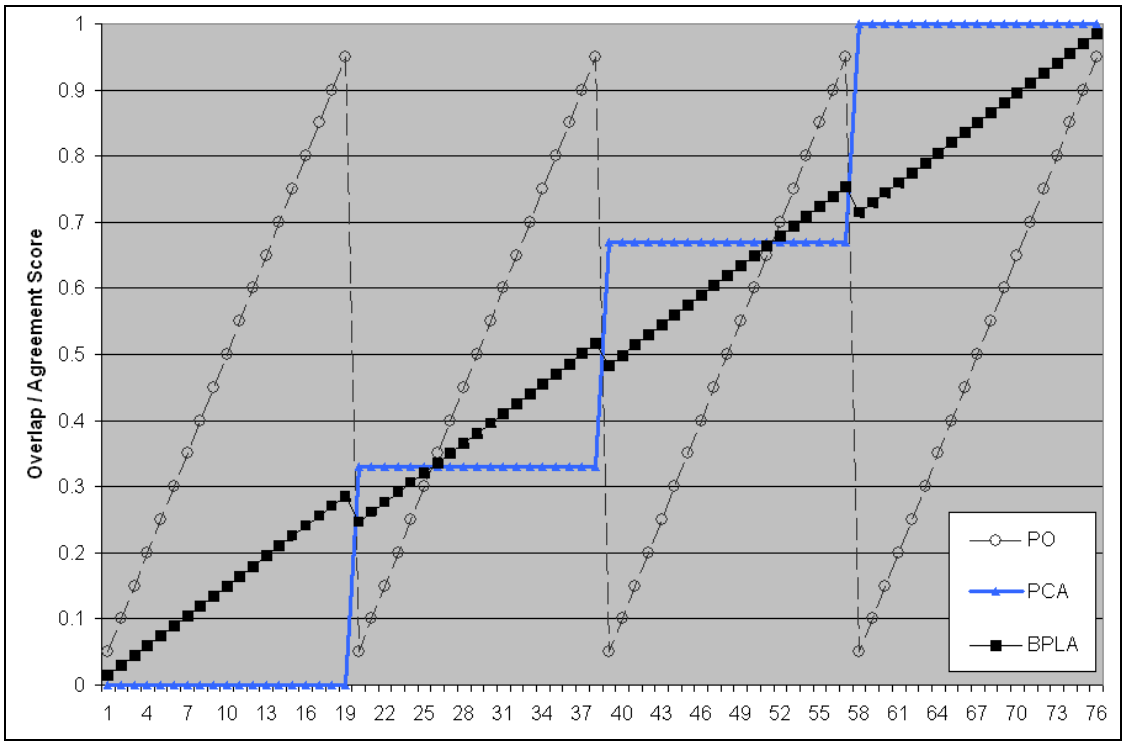

Figure 5.2 Iterations of Bayesian Prior Likelihood Agreement (BPLA) metric for all combinations of Probability Overlap (PO) and Percent Category Agreement (PCA).

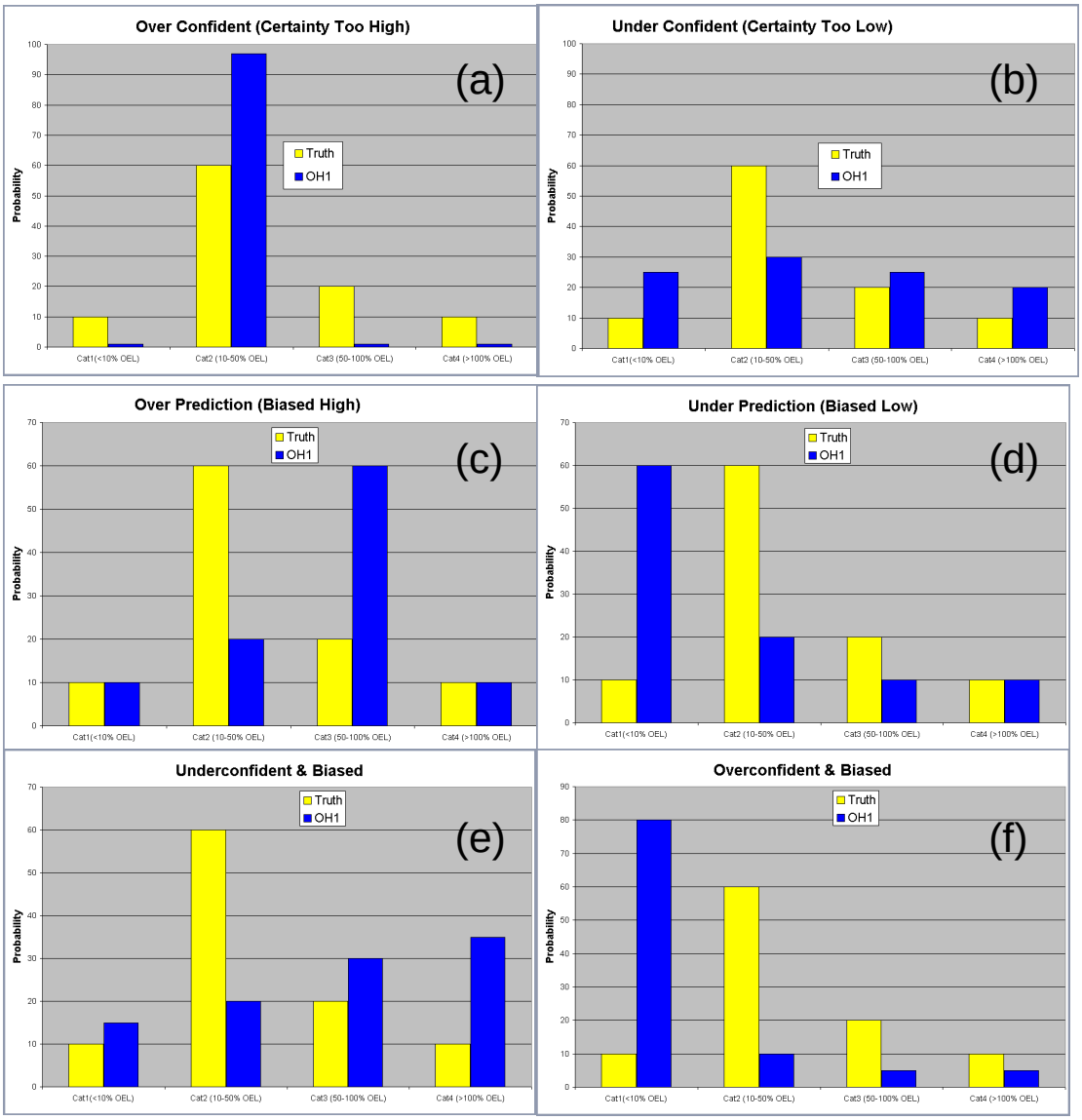

Figure 5.3. Charts of exposure judgments (OH1) versus "Truth" illustrating various possible judgment biases including (a) Overconfidence, (b) Underconfidence, (c) Over prediction / Biased High, (d) Under prediction / Biased Low, (e) Underconfidence with Bias, (f) Overconfident with Bias.

Table V.I. Calculation of Fractional Overlap (FO), Probability Overlap (PO), Percent Category Agreement (PCA), and Bayesian Prior Likelihood Agreement (BPLA) metrics for exposure judgment charts from Figures 5.1 and 5.3.

Prior Likelihood Comparison	Fractional Overlap (FO)	Probability Overlap (PO)	Percent Category Agreement (PCA)	Bayesian Prior Likelihood Agreement (BPLA)
Figure 5.1a	1.00	1.00	1.00	1.00
Figure 5.1b	-0.84	0.00	0.00	0.00
Figure 5.3a	0.81	0.61	1.00	0.88
Figure 5.3b	0.88	0.69	1.00	0.91
Figure 5.3c	0.68	0.58	0.67	0.64
Figure 5.3d	0.58	0.48	0.67	0.61
Figure 5.3e	0.77	0.58	0.33	0.41
Figure 5.3f	0.24	0.27	0.67	0.55

The relationship between the PO and PCA metrics with the BPLA metric can be illustrated by plotting all three metrics for a wide range of combinations (Figure 5.2). By incorporating the amount of overlap, the BPLA metric can help provide feedback when judgments are significantly over or under confident as seen in other studies on professional judgment (Gilovich et at, 2002). Several examples comparing hypothetical exposure judgment charts with Bayesian likelihood charts illustrate how each metric represents the concordance in judgments (Table V.I and Figures 5.1 and 5.2). In general, the BPLA metric appears to correlate much better with judgment agreement than any of the other metrics proposed early on in this study (Table V.I, Figure 5.2)

Additional work will also be needed to further identify conscious and unconscious mechanisms to significantly improve quantitative judgments and dramatically improve qualitative judgments. Deliberate practice of critical cognitive conscious decision making skills can help translate them into more intuitive decision making processes. Top performers in many fields are known to make less errors and faster decisions through thousands of hours of performance and training (Simon, 1973; Klein, 1998; Gilovich et al, 2002; Schneider et al, 2003, Montgomery et al, 2005)). The elements of deliberate practice used by professionals in many fields contain the following elements which likely apply to the science of exposure assessment (Ericsson 2006):

- Engagement of focused training for the purpose of improved performance
- Intrinsic motivation to consistently engage in focused training over a long period
- Practice tasks that are on the upper edge of a practitioner's ability
- Feedback mechanisms that provide knowledge of results

- High levels of repetition

These elements should be very applicable to the field of occupational hygiene and in particular exposure assessment.

The concept of deliberate practice seems utterly obvious when looking at athletes or musicians but may not seem as intuitive in many fields of science. Many studies looking at cognitive activities and many areas of science have recently shown that the concepts underpinning "deliberate practice" apply equally well (Ericsson et al, 2006; Gilovich et al, 2002). A few elements seem to be of critical importance for practicing exposure assessors in moving the body of knowledge for exposure assessment forward.

- Well developed understanding of chemical behavior and the physics of fluids and particles.
- Toxicological principles necessary to properly interpret and assign exposure limits.
- Working knowledge of probability statistics and mathematical foundation for properly using and developing exposure models.

Exposure assessment professionals are positioned well to develop processes and tools to strengthen exposure judgments with the available modeling and Bayesian tools. In particular, tools and training which provide constant feedback on the quality of exposure judgments are critical to strengthening professional competency. Exposure assessment workshops using real exposure scenarios, sampling data and modeling tools should become part of the routine training and calibration exercises used by all exposure assessment professionals.

REFERENCES

- ACGIH American Conference of Governmental Industrial Hygienists (2005). Threshold Limit Values and Biological Exposure Indices.
- American Board of Industrial Hygiene (ABIH) 6015 West St. Joseph, Suite 102, Lansing, MI 48917-3980 http://www.ABIH.org/
- Baron J, (2000) Thinking and Deciding, 3rd Edition, Cambridge University Press, ISBN 0-521-65972-8
- Board of Certified Safety Professionals (BCSP) 208 Burwash Avenue Savoy, Illinois USA 61874-9510 http://www.bcsp.org/csp
- Burstyn I., H. Kromhout / G. Ramachandran, Letters to the Editor, Annals of Occupational Hygiene 46(4): 429-432 (2002)
- CEN (Comité Européen de Normalisation): Workplace atmospheres - Guidance for the assessment of exposure by inhalation of chemical agents for comparison with limit values and measurement strategy. European Standard EN 689, effective no later than Aug 1995 (English version) (Feb 1995).
- Cherrie, J.W. and Schneider, T. (1999). Validation of a new method for structured subjective assessment of past concentrations. Annals of Occupational Hygiene, 43:235-245.
- Cock, J., Kromhout, H., Heederik, D. and Burema, J. (1996): Experts' subjective assessment of pesticide exposure in fruit growing. Scand J Work Environ Health 22:425 432.
- Connolly T., H.R. Arkes, and K.R. Hammond: *Judgment and Decision Making: An Interdisciplinary Reader*. Cambridge, United Kingdom: Cambridge University Press, 2000.
- Corn, M. and Esmen, N.A.: Workplace exposure zones for classification of employee exposures to physical and chemical agents. American Industrial Hygiene Association Journal 40:47-57 (1979).
- Corn, M., and Esmen, N.E. (1979). Workplace exposure zones for classification of employee exposures to physical and chemical agents. American Industrial Hygiene Association Journal, 40:47-57
- Ericsson, K.A., N. Charness, P.J. Feltovich, and R.R. Hoffman: *The Cambridge Handbook of Expertise and Expert Performance*. New York, NY: Cambridge University Press, 2006.
- Ericsson, K.A., R.T. Krampe, and C. Tesch-Romer: The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review.* 100(3):363-406 (1993).
- Friesen M.C., P.A. Demers, J.J. Spinelli, and N.D. Le: Validation of a Semi-Quantitative Job Exposure Matrix at a Soderberg Aluminum Smelter. *Annals of Occupational Hygiene* 47(6):477–484 (2003).

- Ganser G., P. Hewett: An Accurate Substitution Method for Analyzing Censored Data, Journal of Occupational and Environmental Hygiene. 7:223-244 (2010)
- Gigerenzer G, Todd PM, ABC Research Group (1999) Simple heuristics that make us smart, Oxford University Press, ISBN 0-19-514381-7
- Gigerenzer G., Hoffrage U., How to improve Bayesian Reasoning without instruction: Frequency Formats, Psychology Review 1995(102)(4)684-704
- Gigerenzer G., Hoffrage U., Overcoming difficulties in Bayesian reasoning: A reply to Lewis and Keren (1999) and Mellers and McGraw (1999), Psychology Review 1999(106)(2):425-430
- Gilovich, T., Griffin, D., Kahneman, D.: Heuristics and Biases The psychology of intuitive judgment, Cambridge University Press, 2002
- Goldstein W.M., and R.M. Hogarth: *Research on Judgment and Decision Making – Currents, Connections, and Controversies*. New York, NY: Cambridge University Press, 1997.
- Griffin, D., Tversky, A.: The Weighing of Evidence and the Determinants of Confidence, Cognitive Psychology 24, 411-435 (1992)
- Hawkins, N.C. and Evans, J.S.: Subjective estimation of toluene exposures: a calibration study of industrial hygienists. Applied Industrial Hygiene 4:61-68 (1989).
- Hewett P., G. Ganser, A Comparison of Several Methods for Analyzing Censored Data, Annals of Occupational Hygiene., Vol. 51, No. 7, pp. 611–632, 2007
- Hewett, P., Logan, P., Mulhausen, J., Ramachandran, G., and Banerjee, S.: Rating Exposure Control using Bayesian Decision Analysis, Journal of Occupational and Environmental Hygiene, 3: 568–581, 2006
- Hewett, P.: A conceptual model for generating random exposures for use in computer simulations. Technical Report No.05-01. Exposure Assessment Solutions, Inc. (www.oesh.com) (2005a).
- Hewett, P.: Equations for calculating exposure management objectives. Technical Report No. 05-02. Exposure Assessment Solutions, Inc. (www.oesh.com) (2005b).
- HSE (Health and Safety Executive): EH40/99 Occupational Exposure Limits 1999. HSE Books, United Kingdom (1999).
- Hubbard DW, (2007) How to measure anything – Finding the value of “Intangibles” in Business, John Wilery & Sons, ISBN 978-0-470-11012-6
- Ignacio, J. and, Bullock, B. (editors): *A Strategy for Assessing and Managing Occupational Exposures*, Third Edition. Fairfax, VA: AIHA Press, 2006.
- .
- Jayjock, M.A.: Uncertainty analysis in the estimation of exposure. American Industrial Hygiene Association Journal 58:380 382, 1997
- Kahneman D., and A. Tversky: *Choices, Values, and Frames*. New York, NY: Cambridge University Press, 2000.

- Kahneman, D., Slovic, P., Tversky, A.: Judgment under uncertainty: Heuristics and Biases, Cambridge University Press, 1982
- Kahneman, D., Tversky, A., On the psychology of prediction, Psychology Review 80(4)1973
- Kahneman, D., Tversky, A.: Prospect Theory: An Analysis of Decision under Risk, Econometrica, XLVII (1979), 263-291
- Keil,C.: Mathematical Models for Estimating Occupational Exposure to Chemicals, AIHA Press, 2000
- Klein, G.: *Sources of Power, How People Make Decisions*. Cambridge, MA: MIT Press, 1998.
- Kromhout, H., Oostendorp, Y., Heederik, D., Boleij, J.S.M.: Agreement between qualitative exposure estimates and quantitative exposure measurements. American Journal of Industrial Medicine, 12:551-562.(1987)
- Kromhout, H., Symanski, E., and Rappaport, S.M.: A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. Annals of Occupational Hygiene 37:253-270 (1993).
- Leidel, N.A., Busch, K.A., Lynch, J.R. (1977). Occupational Exposure Sampling Strategy Manual. National Institute for Occupational Safety and Health (NIOSH) Publication No. 77-173 (available from the National Technical Information Service (NTIS) Publication No. PB274792.
- Logan P., G. Ramachandran, J. Mulhausen, and P. Hewett: Occupational Exposure Decisions: Can Limited Data Interpretation Training Help Improve Accuracy? *Annals of Occupational Hygiene*, 1-14 (2009)
- Lyles, R.H. and Kupper, L.L.: On strategies for comparing occupational exposure data to limits. American Industrial Hygiene Association Journal 57:6-15 (1996).
- Lyles, R.H., Kupper, L.L., and Rappaport, S.M.: A Lognormal Distribution-based Exposure Assessment Method for Unbalanced Data. Annals of Occupational Hygiene 41:63-76 (1997).
- Macaluso, M., Delzell, E., Rose, V., Perkins, J., Oestenstad, K. (1993). Inter-rate agreement in the assessment of solvent exposure at a car assembly plant. American Industrial Hygiene Association Journal, 54:351-359
- Maidment, S.C.: Occupational Hygiene Considerations in the Development of a Structured Approach to Select Chemical Control Strategies. *Annals of Occupational Hygiene* 42:391-400 (1998).
- Meyer MA, Booker JM, (2001) Eliciting and Analyzing Expert Judgment – A practical guide, Society for American Statistical Association - Industrial and Applied Mathematics (ASA-SIAM), ISBN 0-89871-474-5
- Montgomery H., R., R Lipshitz., and B. Brehmer: *How Professionals Make Decisions*. Boca Raton, FL: CRC Press, 2005.

- Mulhausen, J. and Damiano, J. (editors): A Strategy for Assessing and Managing Occupational Exposures, Second Edition. Fairfax, VA: American Industrial Hygiene Association (1998).
- Nicas M.:An Impractical Emphasis, The Synergist, Sept 2003
- Nicas, M. and Jayjock, M. (2002). Uncertainty in Exposure Estimates Made by Modeling Versus Monitoring. American Industrial Hygiene Association Journal, 63:275-283
- Nicas, M. and Spear, R.C.: A Probability Model for Assessing Exposure Among Respirator Wearers. Part II- Overexposure to Chronic Versus Acute Toxicants. American Industrial Hygiene Association Journal 53:419-426 (1992).
- Otway H., D. von Winterfeld, Expert judgment in risk analysis and management: process, context and pitfalls, Risk Analysis Vol 12 No 1, 1992
- Plous, S, (1993) The psychology of judgment and decision making, McGraw-Hill, ISBN 0-07-050477-6
- Post, W., Kromhout, H., Heederik, D., Noy, D., Duijzentkunst, R.S. (1991). Semiquantitative estimates of exposure to methylene chloride and styrene: The influence of quantitative exposure data. Appl. Occup. Environ. Hyg., 6:197-204.
- Principles of Biostatistics, Marcello Pagano, Kimberlee Gauvreau, Duxbury Press, Belmont, California, 1993
- Ramachandran, G, Toward Better Exposure Assessment Strategies—The New NIOSH Initiative, Annals of Occupational Hygiene, Vol. 52, No. 5, pp. 297–301, 2008
- Ramachandran, G. (2001). Retrospective exposure assessment using Bayesian methods. Annals of Occupational Hygiene, 45(8): 651-667.
- Ramachandran, G., and Vincent, J.H. (1999). A Bayesian approach to retrospective exposure assessment. Applied Occupational and Environmental Hygiene. 14(8):547-55
- Ramachandran, G., Banerjee, S., and Vincent, J.H. (2003). Expert judgment and occupational hygiene: Application to aerosol speciation in the nickel primary production industry. Annals of Occupational Hygiene, 47:461-475.
- Ramachandran. G. (2008). Toward Better Exposure Assessment Strategies - The New NIOSH Initiative. *Annals of Occupational Hygiene.* 52, 297-301.
- Rock J.C., Can Professional Judgment be Quantified?, American Industrial Hygiene Association Journal 47(ppA370) 1986
- Schneider S. L., and J. S. Shanteau: *Emerging Perspectives on Judgment and Decision Research, Cambridge Series on Judgment and Decision Making*. Cambridge, United Kingdom: Cambridge University Press, 2003.
- Simon, H. A., & Chase, W G. (1973). Skill in chess. *American Scientist,61,* 394-403

- Sprenger, A., Dougherty, M., Differences between probability and frequency judgments: The role of individual differences in working memory capacity, Organizational Behavior and Human Decision Processes 99 (2006) 202–211
- Surowiecki, J, (2005) The Wisdom of Crowds, New York, USA, Random House Anchor Books, ISBN 0-385-72170-6
- Teschke K., C. Hertzman, H. Dimich-Ward, A. Ostry, J. Blair, R. Hershler, A comparison of exposure estimates by worker raters and industrial hygienists, Scand J Work Environ Health 1989:15-424-429
- Tuggle, R.M.: The NIOSH decision scheme. American Industrial Hygiene Association Journal 42:493-498 (1981).
- Vadali, Monika, Ramachandran, Gurumurthy and Mulhausen, John, Exposure Modeling in Occupational Hygiene Decision Making, Journal of Occupational and Environmental Hygiene, (2009)6:6,353 — 362
- Walker, K.D., Catalano, P., Hammitt, J.K., Evans, J.S. 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, 2003
- Walker, K.D., Macintosh, D., and Evans, J.S.: Use of expert judgment in exposure assessment: Part I. Characterization of personal exposure to benzene. Journal of Exposure Analysis and Environmental Epidemiology,, 11: 308– 322, 2001.
- Ward, P. Nicola J. Hodges, Janet L. Starkes and A. Mark Williams, The road to excellence: deliberate practice and the development of expertise, High Ability Studies, Vol. 18, No. 2, December 2007, pp. 119–153
- Wild, P., Sauleau, E.A., Bourgkard, E., and Moulin, J.J.: Combining expert ratings and exposure measurements: a random effect paradigm. Annals of Occupational Hygiene 46:479 487 (2002)
- Zhu L.,Gigerenzer G.: Children can solve Bayesian problems: the role of representation in mental computation, Cognition 98(2006) 287-308
- Zsambok C. E., and G. Klein: *Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc, 1997.

APPENDIX 1 - Data Interpretation Training

The data interpretation training consisted of a short (<30 minutes; <40 slides) overview, in which it was suggested that (a) the lognormal distribution is an appropriate statistical model for occupational exposure data, (b) the 95th percentile exposure is an appropriate decision statistic for the exposure limits used in the exercises, and (c) the 95% upper confidence limit (UCL) for the sample 95th percentile can be used to determine with high confidence the range in which the true 95th percentile might lie. The last component of the training consisted of an introduction to (a) several rules of thumb for rapidly estimating an approximate low, middle, and high estimate of the 95th percentile without the aid of a calculator or computer program, and (b) guidance on using these estimates to pick the AIHA exposure control category that most likely contains the true 95th percentile.

The following rules of thumb for estimating the 95th percentile were presented:

- If n is small (e.g., $\leq$6) and one or more measurements exceeds the OEL, then the exposure rating should be Category 4.
- Otherwise, estimate the median exposure and use it as a surrogate of the sample geometric mean (GM): sort the data and determine the median (the median is the middle value if n is odd and the average of two middle values if n is even).
- Multiply the median by three multipliers: 2, 4, and 6.

The results comprise an approximate low, middle, and high estimate of true 95th percentile. The objective is to compare the low, middle, and high estimates to the AIHA exposure control categories with a goal of picking the category that most likely contains the true 95th percentile.

The basis for the three multipliers is the relationship between the true GM and true 95th percentile for a lognormal distribution. The true GM is the median of lognormal distribution. If the true geometric standard deviation (GSD) is low (e.g., 1.5), the true 95th percentile will be approximately twice the GM. If the true GSD is large (e.g., 3), the true 95th percentile will be approximately six times the GM. For intermediate GSD's (e.g., 2 and 2.5), the true 95th percentile will be roughly four times the GM. These rules of thumb use the sample median as a surrogate estimate of the GM.

During the training on the rules of thumb, additional guidance was offered: (1) if the variability within the data set is small (e.g., the low and high values differ by no more than a factor or 2 or 3), then the focus should be on the location of the low and medium estimates of the 95th percentile; and (2) if the variability within the data set is large (e.g., the low and high values differ by a factor that approaches or exceeds 10), then the focus should be on the location of the medium and high estimates of the 95th percentile.

The following eight data sets were used in one of the pre-training data interpretation test exercises (OEL=100 ppm):

A: X = {2, 5, 10, 11, 13, 34}

B: X = {8}

C: X = {9, 18, 24, 43}

D: X = {82}

E: X = {1, 1, 2, 5}

F: X = {2, 11, 26, 35, 60, 118}

G: X = {6, 11, 28}

H: X = {9, 15, 19, 23, 36, 54}

The participants were initially required to inspect the data and decide, using their own set of decision rules, which AIHA exposure control category most likely contains the true 95th percentile. Presumably, at the post-training stage of the study, the participants would be inclined to use the statistical training and rules of thumb, although there was no requirement that they do so. The *correct or "reference"* exposure category was assumed to be the most likely category determined using Bayesian Decision Analysis.

The following table contains the rule-of-thumb calculations for the above data sets, as well as the assigned exposure category determined by majority vote by one of the participant groups at the end of the statistics training. In this instance, the exposure category chosen by the participants matched the category selected using Bayesian Decision Analysis for all eight data sets, demonstrating that the rules of thumb are reasonably accurate when used to select the AIHA exposure category. Note that the purpose of the training was not to suggest that these rules of thumb be used for all future data analysis and interpretation, but instead was to provide the participants, within the context of the study, access to a common distributional model, decision statistic, and set of decision rules.

Data Set	Median	95th Percentile			Exposure Category
		Low	Medium	High	
A	10.5	21	42	63	2
B	8	16	32	48	2
C	21	42	84	126	3
D	82	164	328	492	4
E	1.5	3	6	9	1
F	30.5	(61)	(122)	(183)	4
G	11	22	44	66	2

H	21	42	84	126	3

According to the rules of thumb, Data Set F should automatically be given a Category 4 exposure rating due to the fact that one of the six measurements exceeded the OEL. Data Sets B and D consisted of a single measurement. An inspector-type decision rule would invariably result in an assigned exposure category of 1 and 3, respectively. Application of the rules of thumb resulted in the selection of the next higher exposure category.

APPENDIX II – Matlab code for Exceedance Fraction = 0.01%, Universe Upper Boundary GSD (Dmax) = 4, Population GSD = 2

```
function y=FinalN1to5E_001GSD2(Nsamp)
Sim=1000
OEL=100;
'Run with 1-6 samples on exceedance fraction of 0.01%, gm=7.59, gsd=2 BoundGSD=4.0'
% Run with 1-6 samples on exceedance fraction of 0.01%, gm=7.59, gsd=2
gm=7.59, gsd=2
mu1=log(gm);
sd1=log(gsd);
% Setup Category 2 High Certainty Prior
prior = [.2 .6 .17 .03];
% Define counters Lcount for Likelihood, Pcount for Posterior
count=0;
L1E001count5=0;L1E001count10=0;L2E001count5=0;L2E001count10=0;L3E001count5=0;L3E001count
10=0; L4E001count5=0;L4E001count10=0;L5E001count5=0;L5E001count10=0;L6E001count5=0;
L6E001count10=0; P1E001count5=0;P1E001count10=0; P2E001count5=0;P2E001count10=0;
P3E001count5=0; P3E001count10=0;P4E001count5=0;P4E001count10=0;
P5E001count5=0;P5E001count10=0;P6E001count5=0;P6E001count10=0;
% Setup universe and categories
Ncat = 4;
Gmin = 0.0005*OEL; Gmax = 5*OEL;
Dmin = 1.05; Dmax = 4.00;
LGmin = log(Gmin); LGmax = log(Gmax);
LDmin = log(Dmin); LDmax = log(Dmax);
A1 = log(0.1)+log(OEL);
A2 = log(0.50)+log(OEL);
A3 = log(OEL);
A = [LGmin A1 A2 A3 LGmax];
sigma = LDmin+(LDmax-LDmin)*rand(Nsamp,1);
mu = zeros(Nsamp,Ncat);
likecat = [0 Nsamp];
likemc = [0 Ncat];
   NSAMPmu=10;
   innerfun = zeros(0, NSAMPmu);
   likcat = zeros(0, Nsamp);
   likmc = zeros(0, Ncat);
' NE001=1'
% k is the # of simulations, m is the # of samples used in likelihood
for k=1:Sim;
   Y1(1)=lognrnd(mu1,sd1);
  for j = 1:Ncat;
   for i = 1:Nsamp;
    mumin = max([A(j)-1.645*sigma(i) LGmin]);
      if j == Ncat;
      mumax = LGmax;
      else
      mumax = min([A(j+1)-1.645*sigma(i) LGmax]);
      end
      mu = mumin+(mumax-mumin)*rand(NSAMPmu,1);
      for k = 1:NSAMPmu;
```

```
      innerfun(k) = prod(lognpdf(Y1, mu(k), sigma(i)));
      end
      likcat(i) = (mumax - mumin)*mean(innerfun);
   end
    likmc(j) = mean((likcat))*(LDmax-LDmin);
  end
  likelihood = likmc/sum(likmc);
  posterior = (prior.*likmc) / sum(prior.*likmc);
  count=count+1;
% Test Likelihood Rules
          if likelihood(4)<0.1;
          L1E001count10=L1E001count10+1;
          end;
          if likelihood(4)<0.05;
          L1E001count5=L1E001count5+1;
          end;
% Test Posterior Rules
          if posterior(4)<0.1;
          P1E001count10=P1E001count10+1;
          end;
          if posterior(4)<0.05;
          P1E001count5=P1E001count5+1;
          end;
end;
L1E001count10, L1E001count5, P1E001count10, P1E001count5
count
' NE001=2'
beep
% k is the # of simulations, m is the # of samples used in likelihood
for k=1:Sim;
    Y1(1)=lognrnd(mu1,sd1);Y1(2)=lognrnd(mu1,sd1);
  for j = 1:Ncat;
   for i = 1:Nsamp;
     mumin = max([A(j)-1.645*sigma(i) LGmin]);
       if j == Ncat;
       mumax = LGmax;
       else
       mumax = min([A(j+1)-1.645*sigma(i) LGmax]);
       end
       mu = mumin+(mumax-mumin)*rand(NSAMPmu,1);
       for k = 1:NSAMPmu;
       innerfun(k) = prod(lognpdf(Y1, mu(k), sigma(i)));
       end
       likcat(i) = (mumax - mumin)*mean(innerfun);
   end
    likmc(j) = mean((likcat))*(LDmax-LDmin);
  end
  likelihood = likmc/sum(likmc);
  posterior = (prior.*likmc) / sum(prior.*likmc);
  count=count+1;
% Test Likelihood Rules
          if likelihood(4)<0.1;
          L2E001count10=L2E001count10+1;
          end;
```

```
        if likelihood(4)<0.05;
        L2E001count5=L2E001count5+1;
        end;
% Test Posterior Rules
        if posterior(4)<0.1;
        P2E001count10=P2E001count10+1;
        end;
        if posterior(4)<0.05;
        P2E001count5=P2E001count5+1;
        end;
end;
L2E001count10
L2E001count5
P2E001count10
P2E001count5
count
' NE001=3'
beep
% k is the # of simulations, m is the # of samples used in likelihood
for k=1:Sim;
   Y1(1)=lognrnd(mu1,sd1);Y1(2)=lognrnd(mu1,sd1);Y1(3)=lognrnd(mu1,sd1);
  for j = 1:Ncat;
   for i = 1:Nsamp;
    mumin = max([A(j)-1.645*sigma(i) LGmin]);
      if j == Ncat;
      mumax = LGmax;
      else
      mumax = min([A(j+1)-1.645*sigma(i) LGmax]);
      end
      mu = mumin+(mumax-mumin)*rand(NSAMPmu,1);
      for k = 1:NSAMPmu;
      innerfun(k) = prod(lognpdf(Y1, mu(k), sigma(i)));
      end
      likcat(i) = (mumax - mumin)*mean(innerfun);
   end
    likmc(j) = mean((likcat))*(LDmax-LDmin);
  end
  likelihood = likmc/sum(likmc);
  posterior = (prior.*likmc) / sum(prior.*likmc);
  count=count+1;
% Test Likelihood Rules
        if likelihood(4)<0.1;
        L3E001count10=L3E001count10+1;
        end;
        if likelihood(4)<0.05;
        L3E001count5=L3E001count5+1;
        end;
% Test Posterior Rules
        if posterior(4)<0.1;
        P3E001count10=P3E001count10+1;
        end;
        if posterior(4)<0.05;
        P3E001count5=P3E001count5+1;
        end;
```

```
end;
L3E001count10
L3E001count5
P3E001count10
P3E001count5
count
' NE001=4'
beep
% k is the # of simulations, m is the # of samples used in likelihood
for k=1:Sim;
   Y1(1)=lognrnd(mu1,sd1);Y1(2)=lognrnd(mu1,sd1);Y1(3)=lognrnd(mu1,sd1);Y1(4)=lognrnd(mu1,sd1);
  for j = 1:Ncat;
   for i = 1:Nsamp;
    mumin = max([A(j)-1.645*sigma(i) LGmin]);
      if j == Ncat;
      mumax = LGmax;
      else
      mumax = min([A(j+1)-1.645*sigma(i) LGmax]);
      end
      mu = mumin+(mumax-mumin)*rand(NSAMPmu,1);
      for k = 1:NSAMPmu;
      innerfun(k) = prod(lognpdf(Y1, mu(k), sigma(i)));
      end
      likcat(i) = (mumax - mumin)*mean(innerfun);
   end
    likmc(j) = mean((likcat))*(LDmax-LDmin);
  end
  likelihood = likmc/sum(likmc);
  posterior = (prior.*likmc) / sum(prior.*likmc);
  count=count+1;
% Test Likelihood Rules
          if likelihood(4)<0.1;
          L4E001count10=L4E001count10+1;
          end;
          if likelihood(4)<0.05;
          L4E001count5=L4E001count5+1;
          end;
% Test Posterior Rules
          if posterior(4)<0.1;
          P4E001count10=P4E001count10+1;
          end;
          if posterior(4)<0.05;
          P4E001count5=P4E001count5+1;
          end;
end;
L4E001count10
L4E001count5
P4E001count10
P4E001count5
count
' NE001=5'
beep
% k is the # of simulations, m is the # of samples used in likelihood
for k=1:Sim;
```

```
Y1(1)=lognrnd(mu1,sd1);Y1(2)=lognrnd(mu1,sd1);Y1(3)=lognrnd(mu1,sd1);Y1(4)=lognrnd(mu1,sd1);Y1
(5)=lognrnd(mu1,sd1);
  for j = 1:Ncat;
   for i = 1:Nsamp;
     mumin = max([A(j)-1.645*sigma(i) LGmin]);
       if j == Ncat;
       mumax = LGmax;
       else
       mumax = min([A(j+1)-1.645*sigma(i) LGmax]);
       end
       mu = mumin+(mumax-mumin)*rand(NSAMPmu,1);
       for k = 1:NSAMPmu;
       innerfun(k) = prod(lognpdf(Y1, mu(k), sigma(i)));
       end
       likcat(i) = (mumax - mumin)*mean(innerfun);
   end
    likmc(j) = mean((likcat))*(LDmax-LDmin);
  end
  likelihood = likmc/sum(likmc);
  posterior = (prior.*likmc) / sum(prior.*likmc);
  count=count+1;
% Test Likelihood Rules
          if likelihood(4)<0.1;
          L5E001count10=L5E001count10+1;
          end;
          if likelihood(4)<0.05;
          L5E001count5=L5E001count5+1;
          end;
% Test Posterior Rules
          if posterior(4)<0.1;
          P5E001count10=P5E001count10+1;
          end;
          if posterior(4)<0.05;
          P5E001count5=P5E001count5+1;
          end;
end;
% Output for counts on Likelihood and Posterior Rules
L5E001count10, L5E001count5, P5E001count10, P5E001count5, Count, beep
```