The Use of Expert Panels in Highway Safety: A Critique

Simon P. Washington *
Professor, Arizona State University
Department of Civil and Environmental Engineering
Tempe, AZ 85287-5306
Phone: 480-965-2220
Fax: 480-965-0557
Email: Simon.Washington@asu.edu

Dominique Lord
Assistant Professor, Texas A&M University
Zachry Department of Civil Engineering
College Station, TX 77843-3136
Voice 979/458-3949 / FAX 979/845-6481
d-lord@tamu.edu

Bagwant Persaud
Professor, Ryerson University
Department of Civil Engineering
Toronto, Ontario, Canada M5B 2K3
Phone: (416) 979-5000 (ext. 6464), Fax: (416) 979-5122,
Email:bpersaud@ryerson.ca

*Corresponding Author

Submission Date: November 14, 2008
Word Count = 6,392

Keywords: highway safety manual, highway safety, expert panels, accident modification factors, safety performance functions, expert opinion

ABSTRACT

Expert panels have been used extensively in the development of the Highway Safety Manual (HSM). These panels have been used to extract research information from highway safety experts—research that is often inconsistent in the literature, is relatively scant, or was not conducted under ideal conditions. While the expert panels have been used to recommend agendas for new and continuing research, their primary role has been in the development of Accident Modification Factors (AMFs)—quantitative relationships between highway safety and various highway safety treatments. Because the expert panels derive quantitative information in a qualitative environment and because their findings have potentially significant impacts on future highway safety investment decisions, the expert panel process needs to be described and critiqued. The need for this review should not come as a surprise, as scientific evaluation of analytical and quantitative and qualitative methods employed in highway is a routine activity employed in the profession. This is the first known written description and critique of the expert panel process, and is intended to serve future professionals wishing to conduct such panels.

This paper first provides a background of the role that expert panels serve, how they have been used, and background on expert panels used in other transportation applications. The next section describes the expert panel review process in detail. The paper translates the qualitative nature of the expert panel review process into quantitative measures, so that a detailed understanding of the process is afforded. Then, important questions surrounding the accuracy and precision of expert panel findings are raised, serving to motivate the critique. Conclusions and Recommendations identify areas of potential improvement to be considered in future expert panels.

BACKGROUND

The NCHRP 17-29 research program focused on the scientific evaluation of an array of crash prediction methodologies and analytical methodologies currently being integrated into mainstream products that will be used to evaluate and improve rural multilane highway safety. Due to the extreme importance of these methods on highway safety and ultimately the motoring public, it is vital to scrutinize and evaluate methodologies employed in the profession. One procedure that has received little attention in the literature that plays a fairly significant role in highway safety is the expert panel. Expert panels have been and are being used extensively in various research programs related to and in support of the Highway Safety Manual (HSM) development, and deserve the careful scrutiny that other methods receive (*1*).

The Accident Modification Factor (AMF) or function is a quantitative measure of safety that is integral to the HSM and to the Interactive Highway Design Model (2). The AMF is a safety performance factor or function that relates the safety of a highway with a specific countermeasure or treatment. The factor is typically used to quantify the safety effect of a discrete safety treatment, such that:

$$AMF_i = {}^{C_A}\!/_{C_B} \quad (1)$$

where AMF_i is the AMF for treatment i, C_A is the count of target crashes after treatment i was installed, and C_B is the count of target crashes before treatment i was installed. Target crashes are defined as crashes that can be materially affected by the treatment (e.g. night time crashes affected by lighting), and is most reliably obtained from a well designed before-after study (although not always)(*3*). Examples of discrete safety treatments are the installation of a guard rail, intersection lighting, and a raised median.

Unlike a factor which specifies a constant safety effect of a discrete safety treatment, a crash modification function (sometimes called a safety performance function) typically relates the safety effect of a treatment to the value of one or more treatment-related measures, such that:

$$AMF_i = f\left(X\right) \quad (2)$$

where X is a critical value for estimating the safety effect. For example, X might be the width of a widened median, the percent slope of a flattened roadway side slope, or the unobstructed sight distance to an intersection.

The general analytical approach that incorporates the AMF (factor or function) makes use of a ‘baseline’ or ‘base’ model which predicts crashes for sites of interest (e.g., rural roads, intersections of rural roads, multi-lane highways, etc.) based on exposure (AADT or VMT) and perhaps one or two other factors (e.g. whether the facility is in an urban or rural area). A calibration factor is applied to correct for differences across regions (e.g., cities, states) so as to make predictions applicable to local jurisdictions. AMFs or functions are then applied to obtain estimates of the reduction in crashes after installation of a particular countermeasure or treatment. In simplified form, the AMF approach is given as:

$$(\rho_j)(AMF_i)C_{Without} = C_{With} \tag{3}$$

where C_{With} is the predicted number of crashes at a site with treatment *i*, $C_{Without}$ is the observed number of crashes at the site without treatment *i*, and ρ_j is calibration factor for region *j*, and AMF_i is the expected safety effect of countermeasure *i*.

There is solid support for the AMF approach, as described in Lyon et al. (*4*): “After detailed examination of the data obtained across several states and time periods, the approach proposed in the IHSDM appears to be a sound and defensible approach for forecasting crashes. The approach offers two considerable technical advantages over conventional ‘full model’ approaches for forecasting crashes. First, the often high intercorrelation of explanatory variables with traffic volumes renders isolation of the safety effects of individual variables difficult at best, leading to inconsistent predictions. The algorithm approach skirts this issue by allowing only traffic volumes to be statistically associated with crashes, and by using AMFs derived independently of the prediction model. Second, corrections for driver populations, weather, environmental, and other factors, which are often hard to capture and are inter-correlated as well as correlated with traffic volume, are treated with a correction factor”.

These conclusions (*4*) arose from a detailed and extensive analysis of the safety effects of various geometric and traffic factors for rural intersections across several states. With additional details provided in a companion study (*5*), the authors identified omitted known variables, omitted unknown variables, site-selection bias, countermeasure-selection bias, poorly measured and surrogate variables, and model functional forms as factors that contribute to the difficulty in estimating suitable ‘full’ regression models that related traffic and geometric features to safety, further bolstering support for base models with AMFs.

The current practice involves the estimation of AMFs through the use of expert panels. The use of expert opinions in transportation safety applications is not new. Dissayanake and Lu (*6*) relied on expert opinions to assess the safety needs of special populations. European experts were convened to determine the most effective and reliable air traffic management system for improving safety, operational, and environment performance (*7*). The required characteristics of cockpit weather information systems for NASA's Aviation Safety Program, including some intelligent transportation system technologies, were derived in part from the collection of expert opinions (8). Fukuoka (*9*) developed a unified reliability and analysis environment for assessing European railway network safety using expert opinions and information about failures. Recently and perhaps of most relevance to this research, Harwood et al. (*10*) and Lord et al. (*11*) developed algorithms intended for use in the Interactive Highway Safety Design Model which include AMFs that are derived from expert opinions. In their approach, point estimates of AMFs were derived from a collection of expert opinions and are used to adjust baseline model predictions to estimate the impact of various treatments—as described previously. Little has been written on the exact procedure used to derive AMFs, and their derivation has not been as of yet scientifically scrutinized (like for example the evaluation of the AMF and base model approach cited previously).

Although the remainder of this paper is focused on the *process* by which experts derive AMFs—there remain significant technical issues with AMF development that are often difficult to cope with in expert panels. These technical issues give rise to criticism regarding the technical derivation of AMFs, not the process by which expert opinions are used to derive AMFs. In other words, even if technical issues were addressed well by a panel of experts there still remains the issue of how to culminate their expert opinions—the central focus of this paper. Two such technical issues are particularly noteworthy regarding the derivation of AMFs, and are mentioned here, mainly to highlight unresolved issues deserving of further research.

The first complex technical issue is one of identifying the appropriate context of AMF application. "Context" refers to the process of accurately identifying when and where a countermeasure is going to be effective, and is non trivial. Part of this issue is related to understanding what crashes are affected by a countermeasure—commonly referred to as target crashes (see *3*). As an example, installing lighting is likely to have a significant effect mainly on night-time crashes. Many studies, however, have not identified target crashes and instead have estimated AMFs based on total crashes. Thus, implicit in an AMF derived from an analysis based on total crashes is the contextual assumption that the ratio of target crashes is the same as in the original study sites. This assumption is questionable, at best, and requires research attention. Another contextual issue relates to other quantitative and qualitative conditions that are required to realize effective safety treatments. As general examples, certain countermeasures may be effective only in urban areas, when speeds are relatively high, when rear-end accidents are above a threshold, or when pedestrian activity is significant. These context-specific attributes of safety countermeasures are difficult but necessary to identify in order to realize the benefits of safety investments.

A second challenging technical issue that arises in the context of AMF development is one of independence or non-independence of countermeasures. AMFs are independent only when their contexts for being effective are non-overlapping; otherwise they are dependent. In other words, the target crashes reduced by some countermeasure A must be a different set of target crashes reduced by countermeasure B in order for their effects to be independent. To illustrate, consider the addition of automated red light enforcement and pavement resurfacing, each with their associated AMFs derived via expert panels. Their effects are independent only if these two countermeasures reduce or eliminate different crashes. In this case, pavement resurfacing improves the stopping ability of vehicles. Red light cameras result in more vehicles trying to stop quickly to avoid a ticket. Thus, the combined safety effect of these two countermeasures is likely to be less than their independent effects—they are dependent. All pairings of safety countermeasures will face the same technical challenge, a formidable technical challenge that is worthy of significant research attention. We now turn from these technical challenges to the expert panel process.

THE EXPERT PANEL PROCESS—STEP BY STEP

Prior to critiquing and assessing the expert panel approach used in many HSM applications (from hereon referred to as the HSM expert panel) it is important to document and describe the process. The following steps illustrate in detail the expert panel review process: 1) Identify expert panelists; 2) Set panel meeting date and prepare supporting panel materials; 3) Conduct expert panel meeting; and 4) Disseminate results. The expert panel process is described in the remainder of this section was observed by several of the authors on numerous expert panels. The discussion is generalized so as to make the process generic, and the qualitative aspects of the process are quantified to facilitate later discussion.

Step 1: Identify Expert Panelists

The expert panel consists of nationally recognized experts in the subject matter of interest. It is extremely important that a substantial number of leading researchers be assembled to conduct the expert panel review. While there is no 'magic' number of experts, a panel that is too small may not represent the collective set of views in the profession, while a panel too large might be unwieldy to manage and reach consensus on AMF factors and functions. A number between 10 and 15 experts appears to be an appropriate range to satisfy the need to be representative and manage the tasks charged to the panel in a reasonable amount of time. A number of specific panel member attributes are needed:

- Representation from experts in analytical methods and experimental design as applied to transportation safety, and preferably to the substantive area of interest,
- Geographical representation, so that the collective experience of the experts present can speak to the safety needs of various national stakeholders groups, including rural, urban, eastern, western, and mountainous regions of the US, and
- Specific subject-matter experience in the substantive area being studied. It should not be surprising that selected experts will typically have authored a disproportionate number of the research studies that are discussed during the expert panel review, while other panelists may serve as 'lesser experts'. This may create a potential conflict of interest, as the group consensus may steer

away from the finding of the panelist. This situation cannot generally be avoided in the current expert panel process.

Step 2: Set Panel Meeting Date and Prepare Supporting Panel Materials

A panel meeting date must be set when all expert panelists can attend. The expert panel can take between 2 and 4 full days of deliberation and so a comfortable meeting room with snacks, amenities (restrooms, phones, and internet access) is essential to support a quality meeting. Meeting minutes are needed and either transcribed from recordings or recorded by a meeting secretary. Also, a computer projector and flip charts are needed to support the decision-making and consensus building process.

A critical and significant undertaking at this stage is the preparation of materials used to support the expert panel review. Typically this task is undertaken by the funded group or team conducting the expert panel review. The essential product of this task is to compile copies and/or summaries of all the completed and relevant research related to the treatments to be discussed by the expert panel. This compilation consists of all seemingly relevant and available peer-reviewed research and research summaries conducted nationally and internationally by treatment (e.g. all peer-reviewed research —sometimes non-peer-reviewed research— on replacing yield with stop signs in rural areas). In many cases surprisingly little peer-reviewed research is available relevant to the objectives of the expert panel. A typical expert panel review will have sufficient time to review and evaluate between 15 and 30 treatments, based on the typical 2 to 4 day panel—a rate of approximately 1 treatment per hour. This treatment list is circulated to the panel experts prior to the compilation of the materials, to make sure that important treatments have not been omitted.

Ideally the binder of relevant research is assembled substantially in advance of the expert panel meeting and is distributed to all panelists for their review prior to the meeting. This binder becomes a pivotal tool in the expert panel review process, and also serves as an important reference prior to, during, and after the expert panel meeting.

Finally, experts are assigned a specific topical section of the binder to read in detail and asked to be prepared to discuss the material during the expert panel meeting—often a topic closely aligned with the panelists' expertise. If there a large number of topics an expert may be assigned a set of treatments to review, whereas a small number of treatments may result in overlap among experts. These experts are expected to summarize the research objectively during the expert panel meeting.

Step 3: Conduct Expert Panel Meeting

By the time the expert panel is convened, a comprehensive list of treatments has been endorsed by the panel and a binder of all relevant and available research has been compiled, summarized, and distributed to the experts for prior review. In addition to their subject-matter experience, all experts should arrive at the expert panel meeting having reviewed the materials.

An agenda is determined and distributed that follows a logical sequence for discussing the treatments by group. For example, roadside treatments may make one group, whereas signing and striping may constitute another. The grouping of course depends on the subject matter. Finally, usually there is some hierarchy assigned to the groups, with 'largest impact' treatments being discussed first, 'large impacts' being discussed second, etc. The hierarchy may be determined by the availability of literature on the subject (assumed to be proportional to its importance, with some notable exceptions), with high importance discussed first; or the speculated magnitude of the treatment effect (often correlated with the importance), with high-magnitude effects discussed first; or by the controversiality of the treatments, with less controversial treatments discussed first. Often these three hierarchies are related to one another (i.e., less importance is associated with less research which is associated with more controversy) and so the decision on which treatments to discuss first usually is not a difficult one.

A relatively unstructured open discussion technique is conducted, with a designated moderator leading the general discussion. Treatments are discussed along with the research results for each treatment. The expert review panel's goal is to derive a 'weighted average' AMF factor or function through interactive and open discussion of relevant research and by assigning relevance weights to all of the relevant research. Although the weights are not explicitly (no numbers assigned) or even objectively determined through this process (e.g. ballots), discussion continues until consensus is reached. Based on samples from HSM expert panels, this process may take anywhere between 20 minutes and 3 hours, but is generally not time-constrained in any way—the panel deliberates until an AMF factor or function is agreed upon. It is important to note that a AMF of 1.0 (no effect whatsoever) and the lack of a suitable AMF are outcomes on which the panel may reach consensus.

Mathematically, one might think of this process as developing the following equation,

$$AMF_i = \sum_{k=1}^{K} \varpi_{i,k} AMF_{i,k} \tag{4}$$

where $CMF_{i,k}$ is the AMF for treatment *i* and expert *k*, and $\varpi_{i,k}$ is the weight exerted on treatment *i* by expert *k*. It is important to note that these weights are implicit and not explicit (hence the use of the word “exerted” as opposed to “assigned”. Suppose there are 12 experts (K = 12), and only 6 of the experts provide opinions on the effectiveness of *Stop* to *Yield* signs at four-way intersections. Then, implicitly 6 of the factors have $\varpi = 0$. If one of the remaining 6 experts has done considerable research herself on this topic, then perhaps her implicit weight will be ½, while the remaining experts will share the remaining half with each having 1/10. It is probably not possible with this procedure to determine the weights of the experts, and certainly no effort has been made to do so in past expert panels as regards HSM development—thus equation (4) is meant merely to serve as a mathematical representation of the implicit expert panel deliberation process.

Equation (4) is similar to an equation one might find in a meta analysis, although a meta analysis is significantly more formalized and based on step-by-step procedures, whereas the expert panel is based on consensus building (*12*). Similar to a meta analysis, a number of important factors are discussed with respect to summarizing the available peer-reviewed research, typically quite systematically:

Factor 1: Relevance of the research to the application being discussed. For example, was the research conducted in an urban environment when a rural treatment is being sought? Was the research conducted on mountainous terrain when flat terrain is the setting of interest? Typically these questions of relevance surround issues of traffic exposure, driving population (e.g. country in which research was conducted), range of conditions examined, and similarity of ‘non-treatment’ traffic controls.

Factor 2: Timeliness of the research. The age of the research and its affect on changes in relevance as regards road users, analysis methods, vehicle safety, and injury reporting thresholds is often relevant for discounting the relevance and weighting of research.

Factor 3: Non ideal conditions of the research design. The research conditions that may lead to incorrect or weak conclusions such as omitted important variables, included irrelevant variables, endogeneity of variables, inappropriate analysis methods, or sampling procedure are discussed, with research studies conducted under non-ideal conditions typically discounted or given lesser weight in panel deliberations.

Factor 4: Sample size and sample representativeness. Studies with large samples typically are given greater weight than studies using small samples, all else being equal. In addition, studies with greater sampling representativeness (heterogeneity) of the population are given greater weight than studies conducted on more limited or biased samples.

Factor 5: Findings and conclusions of the research. The conclusions of research are often viewed to make sure the expert panel arrives at the same conclusions as the study authors. While some of the previously listed issues may attract greater attention, studies where authors over- or mis-stated the conclusions are scrutinized.

Factor 6: Consensus on research. Research that confirms prior research, or that represents a substantial body of research that has reached consensus on a topic is more convincing that the lone study. Of course research quality is important here, but assuming equal quality, consensus on the effect of a treatment tends to lend relatively greater credibility.

The expert panel systematically discusses these six factors (and perhaps others) of relevant research for each treatment discussed. Mathematically, this process can be thought to add complexity behind the development of equation (4), whereas the weights “exerted” by experts are the function of their opinions on the factors as they pertain to past research and their experiences with the treatment being discussed, such that:

$$\varpi_{i,k} = \sum_{l=1}^{L} \omega_{i,k,l} Factor_{i,l} \tag{5}$$

where $\omega_{i,k,l}$ is the weight implicitly assigned by reviewer *k* for treatment *i* for *Factor l*, where Factor is one of the 6 (or more) factors described previously. Thus, for example, an expert *k* might implicitly exert equal weights to all 6 of the factors—suggesting that the past research scores very well on all measures for treatment *i* and thus the expert feels quite confident that his opinion of the AMF should receive large weight. Again, these assessments on the factors that influence an experts' weight in the final AMF are unobserved, or latent. However, it is important to conceptualize the process and understand that these implicit evaluations underlie the development of AMFs in the expert panel process.

The session recorder takes notes, records, and otherwise keeps track of conclusions that are drawn regarding all of the AMF and AMF functions as a result of this process. All of the details necessary to derive a AMF factor or function are decided during this meeting, such as the value of the factor, the limits of the function, the shape of the function, and any non-linearities, spikes, or discontinuities. In the majority of cases a computer and computer projector are used so the AMFs can be shown during the meeting and revised to reflect consensus.

Step 4: Disseminate Results

The results of the session are distributed to panel members for review and comment. This final step is conducted to make sure all events and decisions made were captured and are reflected accurately in the AMF factors and functions. After panel members have provided comment, AMFs are described and detailed in a document intended for broader dissemination. The implicit weights and factors that underlie the development of the AMF factors and functions are typically not recorded or documented. In contrast, a meta-analysis would typically document the weights and factors used to derive overall weighted averages.

CRITIQUE OF EXPERT PANEL PROCESS

With theoretical support for the AMF analytical approach and an established history of appropriate uses of expert opinions and panels, the use of expert panels is likely to continue into the near future. There remain, however, some important questions that need to be addressed regarding the derivation of AMFs via expert panels:

1. Are the results derived from expert panels precise and/or accurate?
2. Can expert panels be used to derive estimates of uncertainty?
3. Do results across expert panels differ, and if so, how?
4. Can expert panels be made to ensure repeatable and accurate results?
5. Should expert panels follow informal procedures (as they have been) or more formal and structured procedures such as the Delphi Method?

These questions, which are meant to assess the scientific credibility and use of expert derived AMFs are now addressed in turn. An attempt is made to identify how deficiencies might be tested and/or addressed, and to quantify the issues whenever possible.

Are the Results Derived from Expert Panels Accurate and Precise?

Precision and accuracy of AMFs via HSM expert panels are difficult to assess. If statistical accuracy and precision criteria are applied, then HSM expert panels would need to be repeated numerous times in order to compute the relevant statistics. It is unlikely that this kind of controlled evaluation will be conducted given the enormous resources that would be required.

The answer to this question hinges upon the repeatability of experts in deriving a AMF factor or function given the relevant literature of information. Research by Melcher et al. (*13*) developed a subjective ranking methodology for culling expert opinions and showed that experts generally agreed with one another in evaluating safety treatment effectiveness, demonstrating inter-rater reliability of the instrument. In subsequent work (*14*), the same methodology was applied by a collection of experts to assess the effectiveness of safety investments at intersections of at grade railroad crossings. The researchers derived the AMFs in the original study (*13*) in a different manner than those derived in HSM expert panels, the differences of which are noteworthy. First, experts derived AMF factors independently of each other, and did not engage in a consensus building exercise. Second, experts were given a random sample of crashes in order to examine the effectiveness of a set of treatments rather than a summary of literature findings. Third, multiple observations across experts were tallied to derive means and variances of the AMFs. Finally, only AMF factors were considered (no AMF functions were derived).

The most obvious practical differences between the HSM expert panel derived AMFs and the Melcher derived AMFs are independence and the explicit estimation of the mean. The verbal and non-verbal interaction that occurs within the HSM expert panel is likely to influence the opinions of some experts. Mathematically, the $\varpi_{i,k}$'s and the $\omega_{i,k}$'s in equation (5) are correlated across *k* in the HSM building expert panel, whereas they remain independent when experts are polled independently. Consider, for example, the case when two experts represent a superior and subordinate relationship outside of the panel, would these two experts be more likely to reach consensus? Another example might be two colleagues who worked on the same study together that is relevant to the treatment discussion. Is a panelist who disagrees with the remainder of the expert panel seen as counter-productive? These questions are meant to illustrate that the HSM expert panel process is subject to social interactions and the collaborative goal of reaching consensus that may potentially lead to bias.

To summarize, it is quite likely that the accuracy of HSM expert panel derived AMFs in the absence of bias is similar to that derived by independent methods. The presence and/or extent of bias among experts remains a potentially sticky point and a prime topic of future research.

Can Expert Panels Be Used to Derive Estimates of Uncertainty?

Because consensus is one of the aims of the expert panel, precision is perhaps over-estimated—if it is measured at all. In other words, independent experts (in the absence of consensus building) are likely to disagree more than experts in a consensus building exercise. It is quite reasonable to speculate—in the absence of significant bias—that the accuracy of the two approaches will be similar—and will be bounded by findings in the literature and represent some notion of a mean, median, or mode of AMFs as represented in the literature. In the Melcher expert panels, precision of AMF estimates is reflected by the degree of agreement among experts, such that:

$$\sigma^2\left(AMF_i\right) = \sum_{k=1}^{K}\left(A\bar{M}F_i - AMF_{i,k}\right)^2 \qquad (6)$$

where $\sigma^2\left(CMF_i\right)$ is the variance of the AMF for treatment *i* and $C\bar{M}F_i$ is the mean AMF for treatment *i*. Large disagreement among experts will result in large variance, whereas close agreement among experts (in the Melcher reported approach) will produce small variance. These variance estimates could then be used to reflect the level of confidence that the experts share regarding the treatments effectiveness.

In contrast, precision of AMFs is not systematically estimated in the HSM expert panel approach, as consensus (i.e. high precision) is an explicit goal. Thus, any systematic estimation of precision in HSM expert panel approach is likely to be biased. Instead, the reliability (i.e. variance) of the treatment is estimated in the same fashion as AMFs, experts are polled as to their opinion of the reliability of the treatment (replace CMF_i in equation (4) with $\sigma\left(CMF_i\right)$).

It may be possible, however, to modify the existing expert panel process to poll experts prior to the consensus building process to derive estimates of uncertainty more objectively. This slight modification, if applied consistently and in a structured way, could be used to develop reliable and objective precision estimates.

Do Results Across Expert Panels Differ, and If So, How?

It is possible that different expert panels would produce different AMF factors or functions. As discussed previously these HSM expert panels are not likely to produce AMFs that are substantially different, mainly because the range of possible results is constrained by the literature. As in all research endeavors, a robust expert panel process should not conclude with one expert panel, but will be improved with future expert panels refining and updating AMFs from previous panels. Thus, the expert panel process and the AMFs factors and functions that result should be continually refined and improved with future expert panels and as new research becomes available.

What Guidance can be Provided to Expert Panels to Ensure Repeatable and Accurate Results?

Structure and formality of expert panel procedures in general yield repeatability—as they are the hallmarks of the scientific method. There is considerable structure already included in the expert panel process as described, yielding what are likely to be accurate AMFs. To improve the process, however, there is room for increased formality, particularly when it comes to developing estimates of AMF precision and to address potential problems that result from group social dynamics and potential bias.

Should Expert Panels Follow Informal Procedures (as they have been) or More Formal Expert Panel Procedures such as the Delphi Method?

The Delphi method makes use of questionnaires in two or more rounds of independent polling of panel experts. A facilitator is used to help reach consensus on a forecast (e.g. a AMF factor or function) so that a group of experts may converge on an accurate answer. An administrator is used to initiative and conduct the process, typically done through mail (email or post), thus eliminating travel and meeting related costs. The Delphi method rests on the following principles:

Structured information flow: Unstructured expert panels suffer from the inclusion of irrelevant information and problems associated with group dynamics. In the Delphi method, the initial contributions from the experts are collected via questionnaires, along with open-ended comments to their answers. The panel facilitator controls the interactions among the participants by summarizing the information anonymously and filtering out irrelevant content. This procedure purportedly avoids many of the negative effects of face-to-face panel discussions and solves the usual problems of group dynamics (*16*, *17*).

Regular feedback: In the Delphi process, participants comment on their forecasts, the responses of other experts, and on the progress of the panel as a whole. There are various opportunities to revise prior statements, and these revisions are done anonymously. These revisions are in contrast to unstructured and interactive group meetings, whereby participants tend to stick to previously stated opinions and often conform too much to the group leader. It is for these reasons that the Delphi method is believed to lead to more accurate and objective forecasts.

Anonymity of the participants. In the Delphi method, all expert panelists maintain anonymity throughout the expert panel review process. Their identity is not revealed even after completion of a final report or product. The anonymity purportedly prevents expert panelists from dominating others in the consensus building process by using their authority or personality, frees panelists (to some extent) from their personal biases, and minimizes the "bandwagon” or "halo effect" as discussed previously. The method allows experts to freely express their opinions, encourages open critique, and the revision of prior judgments given the current group consensus.

The Delphi method is thought to extract reliable information from structured groups because of its reliance on feedback in the iterative process of eliciting information from experts (*18*). The Delphi method for polling experts has been shown to produce forecasts that are more accurate than those obtained from unstructured groups of experts (*15*, *16*, *17*), and are intended to pre-empt the kinds of social/psychological/political difficulties that have been found to hinder effective communication and behavior in interactive groups (*16*). For example, unstructured groups are more subject to reaching consensus on issues due to social pressures to conform, and desirability to make progress (*16*). Ayton et al. (*19*) suggests that when high status individuals may be present in an interactive group setting, “people can be characterized as ‘cognitive misers’, and will tend to adopt the simplest coping strategy whenever possible”—one that high status individuals will agree with. Delphi groups, in contrast, have shown significantly less reduction in disagreement across iterations than unstructured groups—suggesting that true disagreement among experts will be reflected in increased variance or uncertainty. When testing has been assessed as regards the accuracy of interactive groups compared to Delphi methods, 5 studies have shown superior accuracy for Delphi methods, 2 studies revealed a tie between the two methods, and 1 study found mixed results (*16*).

Whether or not the expert panel process used in support of the current HSM morphs into a process akin to the Delphi method remains a topic of future study for users of expert panels. It is clear, however, that the Delphi method has been shown to produce more objective forecasts than unstructured panels, is less subject to 'forced consensus', and is less expensive. Future expert panels used to illicit highway safety AMFs and other expert forecasts are well-advised to consider the positive attributes of the Delphi method when considering possible modifications to the process.

CONCLUSIONS AND RECOMMENDATIONS

Upon careful documentation and evaluation of the HSM expert panel process, the following general conclusions are drawn:

The current HSM expert panel process, with all its strengths and weaknesses, is being consistently applied. Consistency is a hallmark of a credible scientific process. In addition, breaking consistency is detrimental to any scientific method. Thus, any changes and/or enhancements made to the current expert panel process should only be considered after completion of the first Edition of the Highway Safety Manual.

The expert panels clearly agree on the mission of the panel—to derive 'the best' and most reliable estimates of AMFs and AMF functions. Persons are selected with this mission in mind and with a track record of conducting scientific research in the subject areas. Thus, although estimates are subjectively derived, all of the participants are intimate with objective procedures for deriving estimates.

It is quite likely that the accuracy of expert panel derived AMFs is quite acceptable and that experts will produce a AMF factor or function that is useful in practice and represents close to a mean, median, or mode AMF factor or function on the subject.

The HSM expert panels do not currently systematically derive precision estimates of AMF factors and functions, and there is a need to have such information. For estimating precision of AMFs, in contrast, expert panels are not as reliable as methods that poll or query experts independently. It is possible to modify the existing expert panel process (post edition 1 of the HSM) to poll experts prior to the consensus building process to derive estimates of uncertainty, or to develop a hybrid Delphi process. This slight modification to current practice, if applied consistently and in a structured way, could be used to develop reliable and objective precision estimates. Any future modifications to HSM expert panels should address how AMF precision should be estimated.

The expert panel process and the AMFs factors and functions that result should be continually refined and improved with future expert panels—as is standard practice for all methods applied in highway safety. The significant technical challenges highlighted previously—AMF context and independence of AMFs—remain as important hurdles in the development of AMFs and will require attention in the research community.

Whether or not the current HSM expert panel process morphs into a process akin to the Delphi method depends upon the goal of future panels and the professional communities' acceptance of the current process. It is clear, however, that the Delphi method has been shown to produce more objective forecasts than unstructured panels. Future expert panels used to illicit highway safety AMFs—after the first edition of the HSM is produced, are well-advised to consider the positive attributes of the Delphi method when considering possible modifications to the process.

The use of the Delphi process would enable the expert panel to avoid a physical meeting which reduces the logistics and cost burden of expert panel meetings considerably.

There are two overall recommendations as a result of the HSM expert panel review. First, the current HSM expert panel process should be revisited upon completion of the first edition of the HSM. No changes should be made to the expert panel process prior to completion of the current HSM edition, and it is believed that the current expert panel process will produce reliable and quite reasonable AMF factors and functions. Existing shortcomings are lack of reliable precision estimates of the AMFs, possible complications arising from interactions and group dynamics, and possible forecasting bias as a result. It may be possible to develop a hybrid expert panel process that utilizes the strengths of the existing HSM expert panel process and the Delphi method.

Second, a comparison of the existing HSM expert panel process and the Delphi method should be conducted. To accomplish this, a panel of experts could be selected and randomly assigned either to the Delphi or HSM expert panel. The two expert panel approaches should be conducted on a limited set of treatments to produce AMF factors and/or functions. The results obtained from these two approaches should be compared, documented, and reported, enabling the research community to quantify the difference, if any, of the two approaches.

REFERENCES

(1) Transportation Research Board, Highway Safety Manual. HSM Web Page: http://www.highwaysafetymanual.org/

(2) US Department of Transportation, Federal Highway Administration. Interactive Highway Safety Design Model: IHSDM Web Page: http://www.tfhrc.gov/safety/ihsdm/ihsdm.htm

(3) Hauer, Ezra. (1997). Observational Before-After Studies in Road Safety. Elsevier Science, LTD. Tarrytown, New York.

(4) Lyon, Craig, J. Oh, B. Persaud, S. Washington, and J. Bared (2003). Empirical Investigation of the IHSDM Accident Prediction Algorithm: Rural Intersections. *Journal of the Transportation Research Board*. National Academy Press, Washington D.C., No 1840. pp 78-86.

(5) Oh, Jutaek, C. Lyon, S. Washington, B. Persaud, and J. Bared (2003). Validation of the FHWA Crash Models for Rural Intersections: Lessons Learned. *Journal of the Transportation Research Board*. National Academy Press, Washington D.C., No 1840. pp 41-49.

(6) Dissayanake, S., and J.J. Lu. Evaluation of Critical Highway Safety Issues of Special Population Groups. *ITE Journal*, V69, 9. 6pp. 1999.

(7) Zografos, K.G. and I.M. Giannouli. Evaluating Cost-Effectiveness of an Air Traffic Management System for Europe: Development and Application of Methodological Framework. *Transportation Research Record 1744*. Transportation Research Board, Washington DC. 2001

(8) Sireli, Y., E. Ozan, P. Kauffmann, S. Gupta, and P. Kachroo. A QFD-Based Decision Model for Application of ITS Technologies to Cockpit Weather Information Systems for General Aviation. *Proceedings of 9th World Congress of Intelligent Transportation Systems*. Chicago, Ill. ITS American, Washington DC. 2002.

(9) Fukuoka, H. Reliability Evaluation Method for the Railway System: A Model for Complicated Dependency. *Quarterly Report of Railway Technical Research Institute*. Tokyo, Japan. 2002

(10) Harwood, D.W., F.M. Council, Hauer, E., Hughes, W.E. and Vogt., A. Prediction of the Expected Safety Performance of Rural Two Lane Highways. Federal Highway Administration Report, FHWA-RD-99-207. Washington, DC. 2000

(11) Lord, D., Geedipally, S.R., Persaud, B.N., Washington, S.P., van Schalkwyk, I., Ivan, J.N., Lyon, C., and Jonsson, T. Methodology for Estimating the Safety Performance of Multilane Rural Highways. NCHRP Web-Only Document 126, National Cooperation Highway Research Program, Washington, DC, 2008.

(12) Kulinskaya, E., S. Morgenthaler, R. Staudte (2008). Meta Analysis: A Guide to Calibrating and Combining Statistical Evidence. Wiley Series in Probability and Statistics. Hoboken, NJ.

(13) Melcher, D., K. Dixon, S. Washington, and E. Wu (2001). "Feasibility of 'Subjective' Engineering Assessments of Road Safety Improvements: Bayesian Development." TRB, National Research Council, Washington, D.C. *Journal of the Transportation Research Board*, No 1758. pp. 36-43.

(14) Bayesian Methodology Incorporating Expert Judgment for Ranking Countermeasure Effectiveness under Uncertainty: Example Applied to At Grade Railroad Crossings in Korea. (2006) *Accident Analysis and Prevention*. Elsevier. Vol. 38 No. 2, pp. 234-247.

(15) Green, K.C., J. Armstrong, and S. Graefe (2007): Methods to Elicit Forecasts from Groups: Delphi and Prediction Markets Compared. *Foresight: The International Journal of Applied Forecasting*, Fall issue.

(16) Rowe, G. and G. Wright (1999): The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, Volume 15, Issue 4, pp. 353-375.

(17) Rowe, G. and G. Wright (2001). Expert Opinions in Forecasting. Role of the Delphi Technique. In: Armstrong (Ed.): *Principles of Forecasting: A Handbook of Researchers and Practitioners*, Boston: Kluwer Academic Publishers.

(18) Rowe, G., and G. Wright (1996). The impact of task characteristics on the performance of structured group forecasting techniques. *International Journal of Forecasting*, Elsevier. Volume 12, pp. 73-89.

(19) Ayton, P., W. Ferrell, and T. Stewart (1999). Commentaries on "The Delphi technique as a forecasting tool: issues and analysis" by Rowe and Wright. *International Journal of Forecasting*, Volume 15, Issue 4, pp. 377-381.