Preprints.org

Article

Not peer-reviewed version

Expert Knowledge Elicitation of Prior Information for Use in Acceptance Sampling and Conformity Analysis

Mike Hicks , Bertrand Colson * , Sam Ellis , Rainer Göb , Steffen Uhlig , John Vandenbemden

Posted Date: 8 August 2025

doi: 10.20944/preprints202508.0638.v1

Keywords: acceptance sampling; lot inspection; prior information; prior distribution

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2025 doi:10.20944/preprints202508.0638.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Expert Knowledge Elicitation of Prior Information for Use in Acceptance Sampling and Conformity Analysis

Mike Hicks [1], Bertrand Colson [2,*], Sam Ellis [3], Rainer Göb [4], Steffen Uhlig [2] and John Vandenbemden [5]

[1] Ministry of Defence, London, SW1A 2HB, UK
[2] QuoData GmbH, Prellerstraße 14, 01309 Dresden, Germany
[3] Ellis & Boscawen Associates Ltd, UK
[4] Würzburg University, Maths department, Emil-Fischer-Straße 40, 97074 Würzburg, Germany
[5] Q-Met-Tech, USA
* Correspondence: colson@quodata.de

Abstract

Prior information can be mobilized in order to develop more economic acceptance sampling plan. The question naturally arises where such prior information may be obtained and whether standardized procedures can ensure reliability and objectivity. Such standardized procedures are often referred to as expert knowledge elicitation, EKE for short. This article reviews the three main EKE protocols in use today, SHELF, IDEA and the Classical Method. A summary and discussion of the differences between these three protocols is then provided. Finally, an example is provided in an annex.

Keywords: acceptance sampling; lot inspection; prior information; prior distribution

---

1. Introduction

Expert Knowledge Elicitation (EKE) is a structured process that can be used to marshal the insights, judgements, and knowledge from a group of experts concerning uncertain facts or values that might arise from a specific field or domain. It is particularly useful in situations where empirical data concerning facts or values of interest is scarce, incomplete, or only available for related cases, and where expert opinion can serve as a critical source of information. EKE is used in fields such as risk assessment, engineering, decision-making, pharmaceuticals, healthcare, environmental science, and policy development, where the subject matter experts' knowledge is essential for modelling, estimating, or forecasting uncertain quantities.

The process involves collecting quantitative information from experts, using formal methods designed to minimize those systematic responses to judgement and decision problems known as cognitive biases (Tversky & Kahneman, 1974), and expressing the uncertainty around this information as a probability distribution. These methods often include structured interviews, remote surveys and/ or group discussions. They are often iterative, involving one or more rounds of value elicitation, feedback, and value adjustment. The goal of EKE is to systematically capture expert judgements, ensuring transparency, reproducibility, and reliability in the decision-making processes.

Effective knowledge elicitation requires careful planning, including the selection of experts, the design of elicitation protocols, and the implementation of procedures that mitigate cognitive biases. By aggregating expert insights EKE provides decision-makers with informed perspectives, particularly when dealing with uncertainty and complex problems.

There are three main EKE protocols in use today, SHELF (Oakley & O'Hagan, The Sheffield Elicitation Framework, 2019), the IDEA (Investigate, Discuss, Estimate and Aggregate) protocol

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2025 doi:10.20944/preprints202508.0638.v1

(Hemming, Burgman, Hanea, McBride, & Wintle, 2018) and the Classical Method (Colson & Cooke, 2018). These protocols differ in the following ways

1. the number of experts that can be involved and the number of values that can be elicited at one time.
2. the use of group discussions and the handling of group dynamics
3. the calibration and/or the weighting of the responses from the experts
4. the handling of cognitive biases
5. the methods used to aggregate and summarize individual responses from the group.

2. The Sheffield Elicitation Framework

2.1. Structured Framework

SHELF is a highly structured and systematic approach which provides detailed guidelines for facilitating expert elicitation sessions. It includes predefined protocols, templates, and tools to train experts to quantify their uncertainty, to elicit their knowledge about the subject and to aggregate the information they provide as a group. These tools help to ensure consistency and rigor in the elicitation process.

2.2. A Group-Based Approach

Although SHELF initially focuses on eliciting individual expert opinions, the final emphasis is on group-based elicitation. A group of between 2 or more experts is convened, usually in a face-to-face workshop environment, trained in the protocol, and then briefed about the subject of the elicitation. Their individual judgements are collected, discussed within the group, adjusted in the light of the discussion, and finally combined using "Behavioural" aggregation, it is the use of this aggregation method which limits the size of the expert group. A consensus distribution is reached by the experts considering what conclusions an independent observer of their discussions might reach. The entire process must be planned thoroughly in advance and the group interactions carefully moderated by the facilitator to ensure a balanced discussion, minimize cognitive bias and to avoid dominance by any individual expert. But it may also be implemented as a solo activity when a decision maker seeks to quantify their own knowledge about an uncertain quantity (O'Hagan, 2019).

2.3. Training Experts in Quantifying Their Uncertainty

SHELF emphasizes the training of experts in uncertainty quantification, recognizing that many experts, while knowledgeable in their domains, may not be familiar with expressing or quantifying their judgements probabilistically or dealing rigorously with uncertainty. The training involves eliciting values for one or more unrelated subject areas, for which experts may have no more knowledge than any other member of the public, e.g. the number of medals that will be won by different countries in future international events or the proportion of uninsured drivers in different American states in different time periods. Different methods of quantifying their uncertainty about these estimates are presented by the facilitator and practiced using these examples. It is important that this training is done as a group immediately prior to the discussion and elicitation for the actual subject matter of interest, irrespective of whether the experts have the viewed the pre-elicitation training material provided by the SHELF protocol.

2.4. Eliciting Probability Distributions

Experts are asked to estimate uncertain quantities by specifying key points of a probability distribution by starting initially with defining the widest feasible range and then narrowing down to the upper and lower quartiles (e.g., 25th and 75th percentiles) before finally considering the median (50th percentile) value. The order these values are elicited are important in reducing cognitive biases

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2025
doi:10.20944/preprints202508.0638.v1

such as Anchoring and Over-confidence. These estimates are then used to construct a distribution that represents the Rational Impartial Observer's (RIO) uncertainty about the quantity in question.

2.5. Facilitator Role

A key feature of the SHELF method is the role of the facilitator. The facilitator plans and guides the process, identifying and selecting experts with relevant knowledge, providing a common set of background material for them and ensuring that they understand the task and adhere to the structured methodology. The facilitator also manages group discussions, encourages the sharing of reasoning behind judgements, and ensures that cognitive biases (such as overconfidence or anchoring) are minimized. The facilitator's assistant will record the discussions and operate the software necessary for fitting probability distributions to elicited values.

2.6. Steps in the SHELF Method:

2.6.1. Preparation and Briefing

The facilitator prepares the elicitation session, including defining the problem, selecting experts, and providing them with the necessary background information.

Experts are briefed on the elicitation process and the specific uncertainties they are being asked to quantify.

2.6.2. Training

This includes calibration exercises to help experts better understand how to express their judgements in probabilistic terms. The training seeks to minimize common cognitive biases, such as anchoring and overconfidence.

2.6.3. Eliciting Individual Judgements

Experts are first asked to make individual estimates about the uncertain quantity, they provide estimates for key points in a probability distribution, the upper and lower feasible bounds, upper and lower quartiles, and a median value. This allows the facilitator to capture each expert's individual perspective before group discussion begins.

2.6.4. Group Discussion

Once individual judgements have been collected, the experts engage in a structured discussion. The facilitator encourages experts to explain the reasoning behind their estimates, highlight any uncertainties, and address any discrepancies between their estimates. The aim is to foster an open exchange of views while avoiding groupthink or undue influence from dominant personalities. Group discussions allow experts to learn from each other and refine their estimates through collective reasoning. This often leads to a more robust and balanced judgement than individual elicitation methods.

2.6.5. Consensus Building

After the discussion, the group works towards achieving a consensus estimate. This involves revisiting the individual estimates in light of the group discussion and, if possible, agreeing on a set of percentiles or a full probability distribution that reflects the collective judgement of the group.

The facilitator plays a crucial role in guiding the consensus process, ensuring that all viewpoints are considered, and that the group's final judgement reflects the full range of uncertainties discussed.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2025 doi:10.20944/preprints202508.0638.v1

2.6.6. Fitting a Distribution

Once the key percentiles or points of the distribution are agreed upon, a probability distribution is fitted to those points using statistical techniques. SHELF provides templates and software tools to help facilitators and experts fit distributions such as normal, log-normal, or other distributions that best represent the experts' estimates.

2.6.7. Documentation and Transparency

SHELF places a strong emphasis on documenting the entire elicitation process. This includes recording the individual estimates, the group discussion, and the rationale behind the final consensus. Although Individual comments are usually anonymized in the documentation this record still helps the decision-makers and external stakeholders to understand how the expert judgements were arrived at and provides a clear audit trail for future reference.

3. IDEA Protocol

The IDEA protocol is a structured and systematic approach to Expert Knowledge Elicitation (EKE), designed to gather expert judgements in a transparent and reproducible manner. IDEA stands for Investigate, Discuss, Estimate, and Aggregate—four key steps that guide the elicitation process.

3.1. A Group-Based Approach

The IDEA method focuses entirely on eliciting individual expert opinions, with the emphasis on mathematically aggregating their final individual responses. The protocol suggests that the best regarded experts in a domain seldom perform as well as expected in group elicitation tasks, so peer recommendations are not used in selecting the group the coordinator instead seeks to assemble a group with as diverse range of experience and knowledge of the subject area as possible. Expert groups of 5 or less can be subject to outliers so the protocol suggests that a group of between 6 and 12 experts is asked, at the first stage, to investigate the subject matter of interest and provide their responses to a pre-prepared set of questions posed by the facilitator and problem owner. These responses are collected and an analyst, after first cleaning and standardizing as necessary (e.g. adjusting for consistent units), compares them to the other responses and their aggregates from the rest of the group and provides this comparison as feedback to the individuals. This first stage often takes place remotely and without sharing the names of the other group members with the individual experts.

3.2. Key Steps in the IDEA Method:

3.2.1. Investigate

In this phase the facilitator presents the experts with a clear and well-defined problem, the available background material and a set of elicitation questions. Such questions would usually conform to the following four-step format:

Question:

> 'What will be the average batch nonconformance rate (X) to produce items 'A' by manufacturer 'B' given that the process is being transferred to a new factory in location C in 2026?'

4-step elicitation:

1. Realistically, what do you think the lowest plausible value for the nonconformance rate, X, will be?
2. Realistically, what do you think the highest plausible value for nonconformance rate, X, will be?
3. Realistically, what is your best guess for nonconformance rate, X?

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2025 doi:10.20944/preprints202508.0638.v1

4. How confident are you that your interval, from lowest to highest, could capture the true value of the nonconformance rate, X? Please enter a number between 50% and 100%

The investigation phase ensures that all experts fully understand the problem, including its scope, assumptions, and context. The shared background information and data help frame the issue, allowing experts to independently form their initial judgements. This first stage helps reduce groupthink and other biases caused by any early interaction among experts.

3.2.2.. Discuss

After the experts have individually investigated the problem, they engage in a structured discussion, usually moderated by the facilitator. The purpose of this phase is to allow experts to share their reasoning, highlight their key assumptions, and discuss any discrepancies in their initial estimates when compared to the group. The discussion is aimed at improving understanding, without forcing consensus. It also helps experts learn from each other and refine their perspectives.

3.2.3. Estimate

Experts are then asked to provide a second round of estimates for the elicitation questions that take account of their discussions. Importantly, the experts still work independently when formulating their estimates to avoid conformity pressure. By eliciting multiple estimates, the IDEA method ensures a diversity of opinions and insights.

3.2.4. Aggregate

The final step involves the analysts aggregating the experts' estimates into a collective judgement. Using a simple mathematical aggregation method such as the mean or median values of the individual responses. The aggregation process provides a synthesized view of expert opinions, presenting the problem-owner or decision-maker with a single value and an indication of the uncertainty around it as provided by the group.

3.3. Documentation and Transparency

IDEA places a strong emphasis on documenting the entire elicitation process. This includes recording the initial and subsequent individual estimates, the individual's performance relative to the group, the discussion of their assumptions and reasoning, and the final mathematical aggregations. This record helps the decision-makers to understand how the aggregated estimates and their associated levels of uncertainty were arrived at and provides a clear audit trail for future reference and reproducibility.

4. The Classical Method

The Classical Method of Expert Knowledge Elicitation (EKE), also known as the Cooke's Method, is a formal structured approach used to aggregate individual expert judgements, particularly in contexts where empirical data is scarce, missing or only available for related cases. The Classical Method focuses on quantifying and weighting expert opinions based on their individual performance, specifically by generating calibration and information scores for each expert. These scores are then used to weight the estimates they give in response to the questions on the subject of the elicitation. The Classical Method minimizes common cognitive biases such as overconfidence and anchoring by only using experts' independent private estimates (there are no group discussions) and by focusing on performance-based weighting rather than relying on consensus-building approaches. The Classical method primarily seeks to make expert elicitation as objective as possible by combining expert judgements in a way that emphasizes the reliability and accuracy of individual experts' performance.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2025 doi:10.20944/preprints202508.0638.v1

4.1. Principles of the Classical Method:

4.1.1. Calibration

Calibration refers to how well an expert's probabilistic estimates align with actual outcomes. In the Classical Method, experts are asked to make predictions on a series of calibration questions—questions for which the true outcomes are known but the experts do not have prior knowledge of them. An expert is considered well-calibrated if, for example, the credible intervals they provide in response to the calibration questions more frequently include the true values. The Calibration score measures the statistical consistency between an expert's predicted probabilities and the observed outcomes. Experts with better calibration scores are considered more reliable in their estimates of unknown quantities.

4.1.2. Information Score

The information score measures the informativeness or precision of an expert's estimates. An expert who provides narrow probability distributions (i.e., estimates with high confidence) would score higher in information than an expert who provides estimates with wider more uncertain distributions. However, this is only valuable if the expert is also well-calibrated—overconfident estimates that are inaccurate are penalized.

4.1.3. Scoring Rules

The Classical Method uses performance-based weighting, meaning that expert judgements are weighted according to how well each expert has performed on the calibration questions. Experts are evaluated using both their calibration and information scores. These two metrics are combined into a weighted score that balances the accuracy (calibration) of the expert's estimates with their informativeness (how precise or uncertain their estimates are). Experts who are both well-calibrated and informative are given greater weight in the aggregation of the final estimates.

4.2. Key Steps in the Classical Method:

4.2.1. Preparation

The facilitator, coordinator and problem owner select as diverse a group of experts as possible, based on their relevant domain knowledge. They must have a deep understanding of the subject matter related to the elicitation problem, if possible, they should be a mix of generalists and specialists and may also come from different fields. As the Classical method is carried out remotely by a questionnaire, without group interactions or multiple rounds, there is no limit on the number of experts that can be included, and a large number of questions can also be included.

A set of questions is prepared for a questionnaire that includes a mixture of both seed and target questions. Seed questions are those with known answers (known to the facilitator/ problem owner and but not to the experts), these will be used to assess the experts' performance. They should be similar in nature to the uncertain quantities of interest, having well-established answers from historical data or scientific evidence. Target questions are the main focus of the elicitation and concern the uncertainties or estimates that require expert judgement. But finding a set of calibration questions where the answers are unknown to the domain experts can be challenging.

4.2.2. Elicitation

The questionnaire should ask the experts to provide their individual probabilistic estimates (e.g., 5th, 50th, and 95th percentiles) for the values of interest in response to all the questions in the questionnaire i.e. for both the calibration questions and the target questions.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2025 doi:10.20944/preprints202508.0638.v1

4.2.3. Aggregation

Each expert is scored based on their performance on the calibration questions, using both their calibration and information scores. The group's expert judgements for the target questions are aggregated using weighted averaging, where the weights are determined by each expert's performance score and combined into a single probability distribution for each target question. This distribution reflects the collective knowledge of the group of experts, with more weight given to those who have demonstrated better calibration and informativeness.

4.3. Documentation and Transparency

The method provides a clear, transparent process for weighting and combining expert judgements. Decision-makers can see exactly how the final aggregated judgements were derived, and which experts contributed the most to the final estimate.

5. Elicitation Methods – A summary of the differences

	SHELF Protocol	**IDEA Protocol**	**Classical Method**
Elicitation Process	Facilitated group discussion aiming for consensus	Independent estimates before and after discussion	Individual estimates with no group discussion
Group size	Small (1 to 6)	Medium (6 to 12)	Large (> 6)
Aggregation Method	Behavioural (consensus-building)	Unweighted mathematical (e.g., mean, median)	Performance-weighted based on calibration
Role of Facilitator	Active, driving consensus	Minimal, process-focused	Minimal, mathematically focused
Focus	Consensus-building	Independent judgements	Individual expert's performance
Discussion and Group Dynamics	Managed to reach consensus	Group discussion to share reasoning, but not to force consensus	No group discussion, focuses on individual estimates
Bias Mitigation	Facilitator-driven, to reduce anchoring and overconfidence	Process driven to reduce anchoring and overconfidence	Bias handled mathematically via calibration and performance weighting
Type of Judgements	Shared probability distributions	Individual probability estimates	Individual probability estimates
Application Context	Regulatory decision-making, policy assessment	Risk analysis, forecasting	Safety, engineering, environmental risk

6. Discussion

6.1. Weighting and aggregation

The effectiveness of the different protocols can be difficult to compare. Any validation of expert elicitations should assess how accurately the experts' beliefs are being represented as well as whether those beliefs represent reality. But as it is only possible to assess the former with an elicitation process, comparisons between the methods usually concentrate on the latter. However, expert elicitations are used to provide estimates for unknown quantities, but for such a comparison it would be necessary to have access to known values that are however unknown to the experts. The classical method partially avoids this problem by providing seed questions related to the original question of interest and with known answers, but which are assumed to be unknown to the experts. Using such a set of questions with each of the different protocols can provide a comparison of their relative accuracy.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2025 doi:10.20944/preprints202508.0638.v1

Some recent work to compare these methods (Williams, Wilson, & Wilson, 2021) has shown that the SHELF method of unweighted behavioural aggregation can achieve better results than those that use performance weighted or other mathematical aggregation techniques.

6.2. Bayesian aggregation

In addition to the *behavioural*, *simple mathematical* and *weighted performance* aggregations as used by SHELF, IDEA and the Classical Method respectively, a Bayesian aggregation method has been proposed (Hartley & French, 2021) that considers the problem-owners current knowledge of the subject area as the prior distribution, the experts' responses as new data, which can be combined through the likelihood function to derive a distribution of the true value given their individual judgements and applies Bayes theorem to derive a final elicited posterior probability.

Appendix A. An Example of an Elicitation Using the SHELF Methodology

Elicitation methodologies typically involve obtaining expert judgements with the assistance of a facilitator. Facilitators can be very effective, although in many applications it is often only a single analyst, without external facilitation, who assess the distributions of data for use in quality assurance. Despite this adopting the facilitator approach by asking questions and challenging responses can help ensure careful judgement and avoid many of the psychological biases associated with deriving such estimates.

The following example illustrates the steps that might be undertaken when using the **SHELF** methodology (Oakley, Eliciting univariate probability distributions, 2010) with one or more experts:

1. **Elicitation Training**
   Before the elicitation, the facilitator asks the experts to review the available training material. This is to ensure that each expert has a consistent understanding of subjective probability and frequency and has had some practise in eliciting parameter values using a number of different methods, e.g. Monte Carlo and bisection methods. On the day of the elicitation the facilitator may again offer an elicitation training session that includes an example elicitation.
   In our example the analyst reviews the freely available Probabilistic judgements e-learning course (New & O'Hagan, 2018).
2. **Review Relevant Evidence:**
   Also before the elicitation and to reduce availability bias, the facilitator ensures that all the pertinent evidence about the product, including perhaps data from other similar manufacturing processes or similar products have been collated. To do this the facilitator should approach each expert asking for them to search for and to submit any relevant information that they have access to. All of the gathered information is then shared amongst all the experts before the elicitation.
   As the product in our example is new, with only limited data from pre-production processing this prompt encourages the analyst to investigate a related process but for a different product the manufacturer already produces.
3. **Define Credible Bounds:**
   On the day of the elicitation and having agreed a suitable method to use, the experts start by specifying their upper (U) and lower (L) credible bounds. Although the data was shared amongst all the experts they will initially be asked to work individually and not share their estimates at this stage with others in the group, this is to reduce possible authority bias. The facilitator should also explain that the bounds are not the absolute theoretical limits but form a credible interval such that there might still be some small probability of it not containing the true value. The facilitator may challenge the experts' estimates by asking how surprised they would be if in the future the true value is considered to be outside of these limits. The experts should, if necessary, adjust their bounds until this scenario appears surprising and unlikely.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2025 doi:10.20944/preprints202508.0638.v1

The credible interval is examined at this stage to ensure all possible values are considered and to reduce the effect of anchoring bias on the subsequent elicitation of the quartiles which can be caused when starting an elicitation with estimating the median.

In our example, from reviewing the related data, the engineer's information concerning the increased complexity of this process, and after some challenges from the facilitator, the analyst estimates that in batches of 1000 manufactured items the lower bound (L) could still be as low as 1 but with 0 being considered unlikely. And that the upper bound (U) could be as high as 100 but a value greater than this would to be implausible.

4. **Specify the Median:**

   The experts estimate their median values, where the true number of nonconforming items in a lot is equally likely to be above or below this value. Although this equal probability judgement appears to be straightforward it is still difficult for experts to do. The facilitator should challenge their estimates by asking each expert to consider which side of their median they would bet contained the true value. The expert should adjust their estimate until they have no preference for either side in such a bet.

   In our example the analyst estimates that the median value (M) is 13 nonconforming items in a lot of 1000 items.

5. **Determine the Quartiles:**

   The experts specify their upper quartiles (Q3) such that the intervals [M, Q3] and [Q3, U] are equally likely, and the lower quartiles (Q1) such that the intervals [L, Q1] and [Q1, M] are also equally likely. This process should ensure, given effective challenges from the facilitator, that all four intervals [L, Q1], [Q1, M], [M, Q3], and [Q3, U] are considered equally probable, each with a 25% probability.

   In our example the analyst estimates Q1 and Q3 to be 5 and 27 respectively

6. **Fit the Distribution:**

   The facilitator fits a distribution to the elicited values (L, Q1, M, Q3, U), ensuring small probabilities outside the credible bounds and equal probabilities across the four intervals. The facilitator then presents the fitted distribution to the expert, offering implied values (e.g., the 5th or 95^{th} percentile) for validation. If the expert disagrees, earlier steps are revisited to refine either the expert's judgements or the fitted distribution until an acceptable result is achieved.

   In our example, the elicitation is being carried out to generate a prior distribution for use in deriving a suitable sampling plan for the quality assurance of the manufacturing process. For the convenience of the analyst and given their previous experience the elicited values were the number of nonconforming items in a manufactured lot of 1000 items. However, as this particular proposal is to be based on sampling by attributes the prior information needs to be in the form of proportions rather than counts. Each is therefore divided by the lot size used in the elicitation to give the values (0.001, 0.005, 0.013, 0.027, 0.100) before fitting a distribution using the SHELF software (Oakley & O'Hagan, The Sheffield Elicitation Framework, 2019)*. The best fit distribution is given by SHELF as a scaled Beta (a = 0.80, b = 3.65, A = 0, B =1).*

7. **Aggregation**

   It is only at this stage that individual estimates should be shared amongst all the experts in the group. A discussion is now encouraged for all to understand any differences and highlight the material referenced in deriving those different estimates. Under the SHELF protocol the individual fitted distributions are not combined mathematically but instead behavioural aggregation is used i.e. having discussed the differences in their estimates the facilitator will ask the group as a whole to agree to new consensus values for the U, Q1, M, Q3 and U values challenging them to include or exclude previous estimates or information depending on outcomes of the discussion phase. The facilitator then presents this fitted distribution to the group, again offering implied values (e.g., the 5th or 95th percentile) for validation. If the group disagrees, earlier steps are revisited to refine either the consensus judgements or the fitted distribution until an acceptable result is achieved.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2025 doi:10.20944/preprints202508.0638.v1

In our example as only one expert was involved in the elicitation this step is not required.

8. **Sample size calculation**

The sample size for the first sample taken from a continuous sequence of lots can be calculated using the elicited prior information concerning the lot conformity, and the producer's and consumer's acceptable risk values.

Producer's risk (PR) = the probability of rejecting (R) the lot when it is conforming (C).

$$P(C\ and\ R) = \sum_{y=Ac+1}^{n} [P(C|y,n,a,b,p_c) \times P(y|n,a,b)]$$

$$PR = \sum_{y=Ac+1}^{n} \left[F_\beta(p_c, a+y, b+n-y) \times \binom{n}{y} \frac{B(a+y, b+n-y)}{B(a,b)}\right]$$

Consumer's risk (CR) = the probability of accepting (A) the lot when it is nonconforming (N).

$$P(N\ and\ A) = \sum_{y=0}^{Ac} [P(N|y,n,a,b,p_c) \times P(y|n,a,b)]$$

$$CR = \sum_{y=0}^{Ac} \left[\left(1 - F_\beta(p_c, a+y, b+n-y)\right) \times \binom{n}{y} \frac{B(a+y, b+n-y)}{B(a,b)}\right]$$

Setting the producer's risk (PR) ≤ 5%, the conforming proportion (p_c) = 0.02, a = 0.8 and b =3.65, from the elicitation, the maximum sample size that meets the producer's risk requirement for an accept zero plan (Ac = 0) is n_p = 65 for PR = 0.05

With the consumer's risk (CR) ≤ 10%, the conforming proportion (p_c) = 0.02, a = 0.8 and b =3.65, from the elicitation, the minimum sample size that meets the consumer's risk requirement for an accept zero plan (Ac =0) is n_c = 23 for CR = 0.096.

The consumer's plan, Ac =0 and n = 23, also meets the producer's risk requirement (as PR = 0.022 ≤ 0.05) and the producer's plan, Ac =0 and n = 65, would also meet the consumer's risk requirement (CR = 0.017 ≤ 0.1). If they didn't then calculate n_p and n_c for the producer's and consumer's plans, for Ac = 1, 2, 3, …, until $n_p \geq n_c$ and therefore any n such that np ≤ n ≤ n_c will meet both risk requirements. Therefore, take n to be the value at the mid-point between n_p and n_c i.e. n = 44 gives PR and CR = 0.038 and 0.037 respectively.

Subsequent sample sizes can be calculated by updating the elicited prior with n, the sample size, and y, the number of nonconforming items in the sample, and repeating the above calculations.

References

Colson, A. R., & Cooke, R. M. (2018). Expert elicitation: Using the classical model to validate experts' judgments. *Review of Environmental Economics and Policy, 12*(1), 113-132.

Hartley, D., & French, S. (2021). A bayesian method for calibration and aggregation of expert judgement. *International Journal of Approximate Reasoning, 130*, 192-225.

Hemming, V., Burgman, M. A., Hanea, A. M., McBride, M. F., & Wintle, B. C. (2018). A practical guide to structured expert elicitation using the IDEA protocol. *Methods in Ecology and Evolution, 9*(1), 169–180.

New, L., & O'Hagan, T. (2018). *Probabilistic judgements e-learning*. Retrieved January 21, 2025, from Sheffield Elicitation Framework e-learning course in probabilistic judgements: https://shelf.sites.sheffield.ac.uk/e-learning-course

Oakley, J. E. (2010). Eliciting univariate probability distributions. In K. Böcker (Ed.), *Rethinking risk measurement and reporting* (Vol. 1). London: Risk Books.

Oakley, J. E., & O'Hagan, A. (2019). *The Sheffield Elicitation Framework*. Retrieved from https://shelf.sites.sheffield.ac.uk/software

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

 doi:10.20944/preprints202508.0638.v1

O'Hagan, A. (2019). Expert knowledge elicitation: Subjective but scientific. *The American Statistician, 73*, 69-81.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124–1131.

Williams, C. J., Wilson, K. J., & Wilson, N. (2021). A Comparison of Prior Elicitation Aggregation Using the Classical Method and SHELF. *Journal of the Royal Statistical Society*(3), 920-940.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.