Expert Judgment in Human Reliability Analysis: Development of User Guidelines

Nicola Paltrinieri[a*] and Knut Øien[a]
[a] SINTEF Technology and Society, Trondheim, Norway

Abstract: Human error probabilities (HEPs) have often been required as input to risk assessments within various industries such as Probabilistic Safety Assessments (PSAs) in the nuclear power industry. In the offshore petroleum industry, however, HRA and corresponding HEPs have only recently been included in some Quantitative Risk Analyses (QRAs). The goal of the Petro-HRA project – for which the work in this paper is part of – is to develop an HRA method for use in the petroleum industry. This will be achieved first of all using the Standardized Plant Analysis Risk Human Reliability Analysis (SPAR-H) method as a starting point. For the adaptation of the SPAR-H method, expert judgment may be needed in the validation and evaluation of the nominal HEP values (a) and the multiplier values (b) of the Performance Shaping Factors (PSFs), if the validation cannot be supported by objective data. For the regular use of the SPAR-H method expert judgment will be needed for the assignment of the PSF multiplier values (c). Each of these three foreseen applications of expert judgment requires the development of specific user guidelines within the Petro-HRA project. The development of the expert judgment user guidelines are presented in this paper.

Keywords: HRA, Expert Judgment, SPAR-H.

1. INTRODUCTION

Human reliability analysis (HRA) is any method by which human reliability is estimated. Human reliability is the probability of successful performance of the human activities necessary for either a reliable or an available system. A realistic assessment of human error probabilities, a major goal of HRA, can reveal weak links in a system that can be corrected before some serious mishap occurs [1].

Human error probabilities (HEPs) have often been required as input to risk assessments within various industries such as Probabilistic Safety Assessments (PSAs) in the nuclear power industry and the space industry [2]. In the offshore petroleum industry, however, HRA and corresponding HEPs have only recently been included in some Quantitative Risk Analyses (QRAs) [3].

The goal of the Petro-HRA project – for which the work in this paper is part of – is to develop an HRA method for use in the petroleum industry [4]. This will be achieved first of all using the Standardized Plant Analysis Risk Human Reliability Analysis (SPAR-H) method as a starting point [5]. This method was created as a simplified HRA method specifically adapted for the Standardized Plant Analysis Risk (SPAR) method used by the US NRC on all nuclear power plants in USA.

For the adaptation of the SPAR-H method, expert judgment may be needed in the validation and evaluation of the nominal HEP values (for diagnosis and action) and the multiplier values (ranges of values) of the Performance Shaping Factors (PSFs), if the validation cannot be supported by objective data. For the regular use of the SPAR-H method expert judgment will be needed for the assignment of the PSF multiplier values. Each of these foreseen applications of expert judgment requires the development of specific user guidelines or expert judgment processes within the Petro-HRA project. The development of the guidelines will be supported by a focused literature review.

The user guidelines are developed for a) validation/evaluation of NHEPs, b) validation/evaluation of PSF multiplier values, and c) rating of PSFs in regular use of the method.

* Nicola.paltrinieri@sintef.no

2. STATE OF THE ART

When dealing with failures in decision or diagnosis tasks, data collection for HRA may be relatively challenging. A number of variables, such as the contextual factors, can increase the sensitivity of related decisions. Moreover, the lack of data can further jeopardize the analysis.

In order to deal with these issues, HRA methods based on Expert Knowledge became popular in the 1980s [6]. SLIM (Success Likelihood Index Methodology) was developed within the U.S. Nuclear Regulatory Commission (NRC) to obtain estimates of HEPs by using expert judgments [7]. The APJ method (Absolute Probability Judgment) is another example of an expert judgment-based approach [8]. Also more recent methodologies focusing on decision and diagnosis tasks mostly rely on expert judgment, as explained in the related guidelines: ATHEANA [9], MDTA [10], CESA [11], MERMOS [12].

In THERP [13], HEART [14], HRA Calculator [15] and SPAR-H [16], expert judgment is combined with empirical data, such as specific HEPs for reference tasks and multipliers to account for different conditions. Finally, Bayesian approaches are suggested by recent publications [17,18] in order to deal with multi-expert knowledge and increased HEP reliability.

2.1. Information on SPAR-H

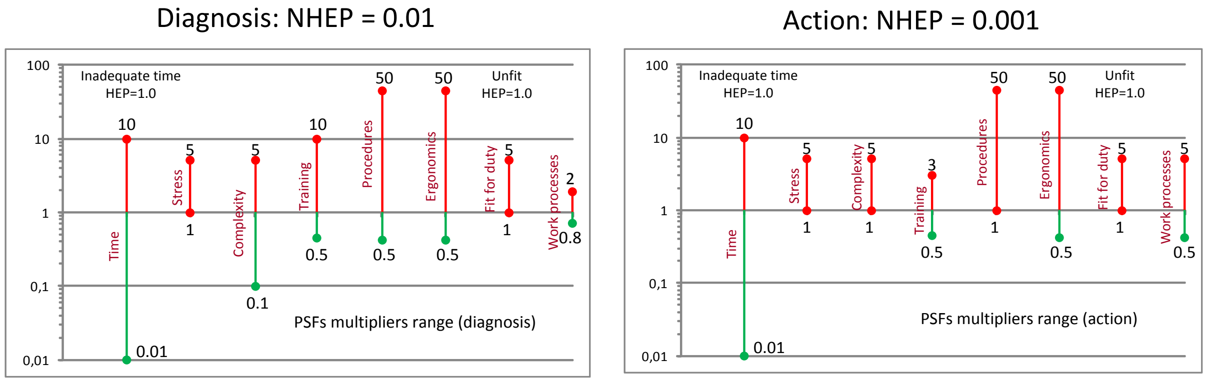
The SPAR-H Step-by-step Guidance [16] describes 5 main steps to fulfill for the application of this technique:

Step-1: Categorizing the HFE as Diagnosis and/or Action
Step-2: Rate the Performance Shaping Factors
Step-3: Calculate PSF-Modified HEP
Step-4: Accounting for Dependence
Step-5: Minimum Value Cutoff

Step-2 "Rating of the PSFs" is the step in which use of structured expert judgment is foreseen. It is about assigning numerical values, although qualitative guidance is provided in rating the PSFs. Eight Performance Shaping Factors (PSFs) are presently accounted for in the SPAR-H quantification process [19]: (i) Available time, (ii) Stress and stressors, (iii) Complexity, (iv) Experience and training, (v) Procedures, (vi) Ergonomics (including the human-machine interface), (vii) Fitness for duty, (viii) Work processes.

These PSFs are rated in Step 2, as described above, and will modify the nominal HEPs in Step 3. The nominal HEPs (for diagnosis and action tasks) and the ranges of the PSF multiplier values are shown in Figure 1 [3].

Figure 1: SPAR-H NHEPs and PSF Multiplier Values (Ranges) (from [3])



2.1.1. Expert Elicitation and SPAR-H

A simplified guideline was also developed within the NRC for seeking expert judgment for risk assessment of operating events [20]. The guideline is structured in two parts: expert elicitation for hardware and for human error. The second part can be taken as an example since it makes use of SPAR-H, as shown in Table 1.

Table 1: Expert Elicitation Guideline for Use in SPAR-H [20]

	Step	Comments
1	Meet Requirements	• There is insufficient information for probability calculation • The event is risk significant • The event is new, rare, complex, or poorly understood
2	Frame Problem	Use predefined worksheet for background information collection [20]
3	Identify Experts	Consult human factors or operations experts
4	Conduct Estimation	Use 2-3 experts and appropriate SPAR-H worksheet [16]
5	Hold Expert Panel	Use predefined worksheet for expert panel [20]
6	Input into Risk Analysis	Arithmetic average of experts' HEPs may represent the analysis input

2.2. Existing Expert Judgment Guidelines

Guidelines for structured expert judgment have been developed by SINTEF for estimation of various types of parameters, such as failure rates and cost data [21,22]. We have used these guidelines as a basis and performed a focused literature review in the further development of the SPAR-H specific expert judgment guidelines. In addition, information on the HRA method to be used has been reviewed and used as a starting-point [5,16].

The process is composed of three distinct phases; preparation, elicitation and calculation, each with a set of procedure steps (Table 2) [21,22].

Table 2: SINTEF Expert Judgment Guidelines – Procedures Steps

	Preparation	Elicitation	Calculation
1	Problem description	Information to the experts	Calibration of experts
2	Evaluation of available resources	Accomplishment of elicitation	Weighting of experts
3	Selection of experts	Establishm. of 3-point estimates	Calculation of comb. estimates
4	Evaluation of experts	Basis for weighting of experts	Documentation
5	Selection of quantification method	Final discussion	Communication/presentation
6	Development of a questionnaire	Possible adjustment of estimates	Supplementary work

The guidelines for preparation, elicitation and calculation are described in a template consisting of procedure steps and examples, pitfalls to be avoided, and references and comments to both the procedure steps and the pitfalls. In addition page number, revision date and revision number are included in the header.

3. EXPERT JUDGMENT PROCESSES FOR ADAPTATION AND USE OF SPAR-H

Some overall principles or constraints to the development of expert judgment user guidelines are:

1. SPAR-H is the starting-point
2. The method needs to be adapted to the petroleum industry
3. The method should be kept simple (and if possible simplified even more)
4. The method should be improved within the constraints given in points 1-3

The two last constraints almost represent a paradox; the method should be improved, but not on the expense of simplicity, and it should be simplified, but not on the expense of the quality of the method.

However, a main reason to keep it simple is that HRA is only a minor part of a PRA within the nuclear industry, and if HRA is going to be successfully implemented in QRAs within the petroleum industry, it needs to be at least as efficient as in the nuclear industry and only constitute a minor part of a QRA.

In the Petro-HRA project, guidelines for structured expert judgment are needed for two reasons:

I. Validation/evaluation of nominal HEPs and the ranges of the PSF values in lack of objective data (including possible combined use of objective data and expert judgment).
II. Use of expert judgment in each application of SPAR-H to determine PSF multiplier values.

While the first process involves a "once-and-for-all" study of the basic parameters to be performed by the project team within the project, the second process relates to the regular use of the SPAR-H method by practitioners after the project. The first process consists of two sub-processes: a) Validation/evaluation of NHEPs, and b) validation/evaluation of PSF multiplier values.

Some basic requirements for the expert judgment process, that have shaped the development of the guidelines, are [3]:

- The expert judgment process needs to be structured
- The expert judgment process needs to be cost-effective
- Experts need to be informed, but without being biased

The expert judgment process needs to be *structured* to obtain a best possible result and to reduce variability between SPAR-H analyses. It should be conducted in a similar manner independent of the analyst.

The expert judgment process needs to be *cost-effective*. In a similar way as HRA provides input to QRA (and only can be a smaller part of the total risk analysis) expert judgment provides input to HRA (and can only be a smaller part of the overall HRA). An exception can be made for the validation of the nominal HEPs and the PSF multiplier values, since this will be a "once-and-for-all" effort.

When using domain experts to provide numerical values or ratings, it is important to prepare them and *inform* them, but *without* influencing or *biasing* their judgment.

4. RESULTS

Expert judgment processes have been developed for the following purposes, in the development/ adaptation and regular use of the Petro-HRA method:

I. a) Validation/evaluation of NHEPs
I. b) Validation/evaluation of PSF multiplier values
II. Rating of PSFs (determining PSF multiplier values) in regular use of the method

The results are presented for each of the three purposes/applications separately. Considerations within each of the three phases: preparation, elicitation and calculation are presented. Note that the expert judgment processes have not been tested or used yet, since the development of the Petro-HRA method is still on-going.

4.1. Expert Judgment Process for Validation/Evaluation of NHEPs

<u>4.1.1. Phase I: The Preparation Phase</u>

The values to be estimated are generic nominal human error probabilities, i.e. values for "a typical" task carried out under nominal circumstances, meaning nominal values of the performance shaping factors.

The current values used in SPAR-H for the nuclear industry are illustrated in Table 3.

Table 3: Nominal Human Error Probabilities (NHEPs) for Diagnosis and Action

NHEP 	Every time	1 in 10	1 in 100	1 in 1000	1 in 10000	1 in 100000
	1.0	0.1	0.01	0.001	0.0001	0.00001
Diagnosis 			✓			
Action 				✓		

Available resources for the expert judgments are related to budget (hours or money), time, statistical basis of data, and number of experts. The budget is based on the budget allocated for validation within the Petro-HRA project. Due to limited resources, the expert elicitation session will be planned as a one-day workshop, with 3-4 experts. A necessary input for this session is the exact basis for the NHEPs currently used in SPAR-H.

The experts could be operators, supervisors, instructors on simulators, or HRA analysts. The experts have not been selected yet. See Section 5 for further discussion – issue 1. Evaluation of experts is relevant if they are to be weighted differently, but due to budget constraints (including few experts), we will give the experts equal weight in the final calculations. The method of calculation will be based on combined individual estimates using a weighting model with or without data. For the NHEPs estimates we will consider using uncertainty estimates.

The questionnaire has been developed based on considerations on contents, appearance/form, how to ask the questions, decomposition of the problem/estimate, the sequence of the questions, testing and training, and the interview format itself. The questionnaire consists of introductory information about the NHEPs, an illustration of what to be estimated, i.e. Figure 3, and a set of questions.

Questions related to the NHEPs are:

1. Are there any reasons why the NHEPs for use in the petroleum industry should differ from the values used within the nuclear industry?
2. What do you think is the NHEP for diagnosis? If it is 0.01 (i.e. 1 in each 100 tasks), mark it with a circle around the green check mark (✓), and if you think it is another probability, mark it with a cross (✖) in the corresponding cell in Figure 3.
3. What do you think is the NHEP for action? If it is 0.001 (i.e. 1 in each 1000 tasks), mark it with a circle around the green check mark (✓), and if you think it is another probability, mark it with a cross (✖) in the corresponding cell in Figure 3.
4. What are your arguments for the judgments?
5. Any other comments?

For the judgments of NHEPs we consider including assessments of uncertainty through the assignment of 3-point estimates.

4.1.2. Phase II: The Elicitation Phase

As an introduction to the elicitation session the experts are informed about the expert judgment process and what to be assessed. It will be explained how the elicitation process will be carried out (including an estimation of the time it will take), how the results will be treated, and how to fill out the questionnaire.

The elicitation will be carried out individually in an elicitation session using the questionnaire, and alternatively followed by a group session at the end. See Section 5 for further discussion – issue 2.

Uncertainty estimates will be considered for the NHEPs by establishing 3-point estimates. In such a case we ask the experts first to indicate the lowest and highest estimates, and finally they indicate their best estimate (the median value). This is to reduce problems with anchoring and overconfidence [23].

We need not establish a basis for giving different weights to the experts, since without a rational basis to weight the experts, unequal weighting should not be used. Weighting leads to a more extensive expert judgment process and is not used in very simple expert judgments [22].

A final discussion can be carried out in a group session at the end to give arguments for the assessments and possibly adjust the estimates.

4.1.3. Phase III: The Calculation Phase

Systematic deviation due to under- or overestimation can be corrected by calibration, based on the use of control questions. Since we do not use control questions for assignment of unequal weights to the experts, calibration will not be carried out in this case.

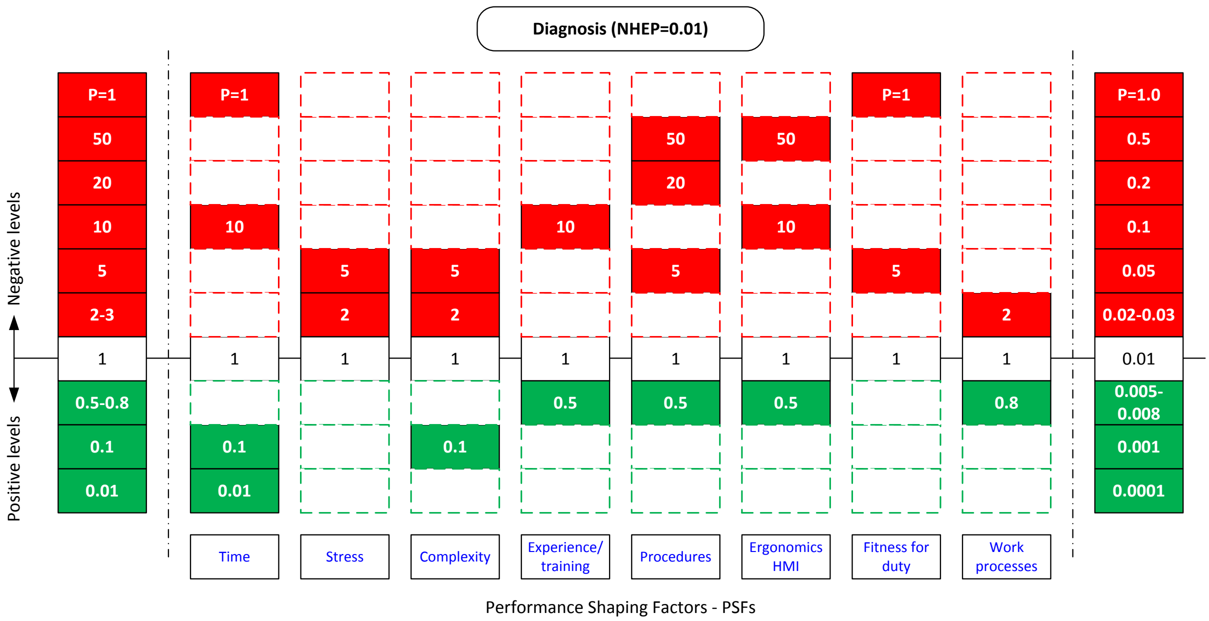
Calculation of a combined estimate is carried out to achieve the estimate that the analyst need for further calculation, decision making, etc. The combined (final) estimate is based on the experts' individual estimates/arguments, whether these are provided individually or in a group. For the NHEPs we will use a classical weighting model with or without statistical data for the calculations.

The expert judgment process will be thoroughly documented and the results communicated to both decision makers and the experts (as feedback to their judgments). This may be followed by supplementary work, e.g. updating of the generic expert judgment guidelines.

4.2 Expert Judgment Process for Validation/Evaluation of PSF Multiplier Values

4.2.1 Phase I: The Preparation Phase

Figure 3: PSF Levels and Values for Diagnosis



The PSF multiplier values correspond to positive or negative effect – compared to a nominal situation – on the probability of human error when performing either a diagnosis task or an action task. The current values used for PSF multiplier values related to diagnosis tasks are shown in Figure 3 (a similar figure applies for action tasks). The figure also shows the levels that are presently not used (dashed boxes). The column to the right shows the modified human error probabilities (given that one PSF on the corresponding level is taken into account).

The evaluation of resources available is similar as for the estimation of NHEPs, whereas the choice of experts may differ. For the estimation of PSF multiplier values it is more likely that experts will be chosen from within the Petro-HRA project group. As for the NHEPs we will not evaluate experts in order to give them unequal weight, due to a limited budget and few experts. The method of calculation will be a classical weighting model without data, since data for PSF multiplier values are most likely non-existent, and only estimates on which levels to choose will be included.

The questionnaire has been developed based on the same considerations as for the NHEPs in Section 4.1.1. The questionnaire consists of introductory information about the PSF multiplier values, an illustration of what to be estimated, i.e. Figure 3, and a set of questions. Questions related to the PSFs multiplier values (and also the PSF levels) are:

1. Is it reasonable that "time" and "fitness for duty" are the only PSFs that in the worst case can increase the HEP by two orders of magnitude and "guarantee" failure?

2. The number of negative levels for each PSF varies between 1 and 3 (i.e. 2-2-2-1-3-2-2-1).
    a. Are there any good reasons why just "procedures" should have 3 levels?
    b. Are there any good reasons why "experience/training" and "work processes" should have only 1 level?
    c. If a reason for having only 1 level for "work processes" is that the maximum effect is a doubling of the nominal HEP (and the maximum positive effect is just 0.8), is it then strictly necessary to include "work processes" as a PSF?
    d. Could we simplify and standardize on 2 negative levels for each PSF, e.g. a "minor negative effect" and a "major negative effect"?

3. The number of positive levels for each PSF varies between 0 and 2 (i.e. 2-0-1-1-1-1-0-1).
    a. Are there any good reasons why just "time" should have 2 levels?
    b. Is it reasonable that "stress" and "fitness for duty" are the only PSFs that may not reduce the human error probability, i.e. they cannot be better than nominal?
    c. The positive effect of "work processes" is only 0.8. Is it unlikely that the effect can be at the same level as other PSFs, i.e. 0.5?
    d. If yes, is it then strictly necessary to include this "very minor positive" change in effect?
    e. Could we simplify and standardize on 1 negative level for each PSF, e.g. a "minor positive effect", corresponding to approximately a half or one order of magnitude decrease in HEP (perhaps with the exception of "time" which also may include a change of two orders of magnitude)?

4. Based on the arguments for the present PSF multiplier values in SPAR-H, the qualitative descriptions, and a comparison across the PSFs, what are your judgments for the multiplier values / levels for each PSF?

Similar questions are asked for PSFs for action tasks.

<u>4.2.2. Phase II: The Elicitation Phase</u>

The elicitation phase will be similar as for the NHEPs, described in Section 4.1.2, with one exception; we will not consider the use of 3-point estimates for the PSF multiplier values.

4.2.3. Phase III: The Calculation Phase

Also the calculation phase will be similar as for the NHEPs, described in Section 4.1.3. The only difference is that we will not consider a calculation model with statistical data, as it is unlikely that we will find data supporting the PSF multiplier values.

4.3. Expert Judgment Process for Rating of PSFs

When the PSFs, their descriptions, levels and multiplier values have been settled, we need to have a structured procedure for rating them in each application of the Petro-HRA method. In the present use of SPAR-H worksheets have been developed for the analyst to document the assigned PSF multiplier values as well as the subsequent calculations of modified HEPs [5]. It is the analyst that assigns the multiplier values based on qualitative information from the "experts".

We suggest extending this approach by adding worksheets for a structured expert judgment process, in which the analyst documents the experts' quantitative assessments, and then transfers the results to the analyst worksheets for further calculations.

4.3.1. Phase I: The Preparation Phase

The values to be estimated are the PSF multiplier values or the PSF levels (with corresponding values).

Available resources for the expert judgments are related to budget (hours or money), time, statistical basis of data, and number of experts. The budget depends on how many tasks we need to assess and provide PSF multiplier values for. At the same time we need to keep the constraints in mind, i.e. that this is only input to the complete HRA, which in turn is only input to the QRA. Finally, the assignment of PSF multiplier values have great impact on the final HEPs, and therefore requires a certain degree of trustworthiness, in particular since the HRA input to the QRA may be met with some skepticism from the risk analysts, due to its novelty in the petroleum industry.

A tentative estimate is that a half day workshop (elicitation session) is needed for one task, and that a one day workshop is needed for more than one task. For "many" tasks, two days may be needed. A tentative first estimate is therefore between half a day and two days. More than two days may be "negotiated" for, if this is essential for a trustworthy result. It is advisable to schedule expert elicitation sessions as early as possible in the HRA process, since the rating of the PSFs is the second step in SPAR-H, as discussed in Section 2.1.

It is not likely that statistical data can be provided for the PSFs as basis for assigning multiplier values. If the experts themselves are able to point to relevant data during their judgments, this should of course be documented as arguments for the judgments.

Preferably we should use minimum two experts, and even check or re-estimate those estimates that differ by a third expert (or by reaching a consensus between the original two experts in a "group" session at the end of the individual estimation session). Alternatively, sensitivity analyses could be performed for those PSFs where the multiplier values differ between the experts. The most suitable expert would probably be an operator carrying out exactly the task in question on an installation in the operation phase, or a supervisor (senior operator) of such operators. See Section 5 for further discussion – issue 3.

Due to keeping the Petro-HRA method simple and efficient, and due to the use of rather few experts it is seen as most appropriate to give the experts equal weight. However, we may need some evaluation/ consideration of the experts with respect to their knowledge about probabilities and multipliers, to decide on whether to focus on the qualitative PSF level descriptions, the multiplier values, or both.

The questionnaire (or expert judgment worksheets) has been developed based on considerations on contents, appearance/form, how to ask the questions, decomposition of the problem/estimate, the sequence of the questions, testing and training, and the interview format itself. The questions to be asked, or how to fill in the questionnaire, will be described on the front page of the questionnaire/ worksheets. We will only ask for "best estimates", since there is very few levels to choose from for each PSF. Uncertainty will rather be reflected by any difference between experts in assigning PSF levels.

The expert judgment worksheets front page is shown in Figure 4 including an example of a PSF worksheet at the bottom. Each worksheet will start with a general description of the PSF on the top of each PSF sheet/page.

Figure 4: Petro-HRA Expert Judgment Worksheets for Assignment of PSF Multiplier Values

Petro-HRA Expert Judgment Worksheets	
Plant/Installation	
Human Failure Event	
ID/Code	
Description	
Scenario	
Expert Judgment	
Facilitator/Analyst	
Expert	
Date	
Special information	
Diagnosis, action or both	
Notes/Comments	

To be filled out by the analyst

If the HFE involves diagnosis, the first 8 sheets (one for each PSF) shall be used; if the HFE involves action, the last 8 sheets (one for each PSF) shall be used; and if both diagnosis and action are involved, all 16 sheets shall be used.

How to fill in the worksheets:
1. Read the descriptions of the PSF levels and decide on the level you judge to be the correct level and put a check mark in the multiplier column
2. Provide arguments for your judgments in the column to the right
3. If relevant, provide additional comments at the bottom

EXAMPLE:

1 2 3

Diagnosis PSF levels				
PSFs	**PSF Levels**	**Multiplier**	**Definition/description/guidance of PSF levels**	**Please note specific reasons for PSF level selection in this column**
Available Time	Inadequate time	P(failure)=1	If the operator cannot diagnose the problem in the amount of time available, no matter what s/he does, then failure is certain.	**In my experience we have extra time available compared to we need, which is approximately two times greater than needed.**
	Barely adequate time	10	2/3 the average time required to diagnose the problem is available.	
	Nominal time	1	On average, there is sufficient time to diagnose the problem.	
	Extra time	0.1 ✓	Time available is between one to two times greater than the nominal time required, and is also greater than 30 minutes.	
	Expansive time	0.01	Time available is greater than two times the nominal time required and is also greater than a minimum time of 30 minutes; there is an inordinate amount of time (a day or more) to diagnose the problem.	
	Insufficient information	1	If you do not have sufficient information to choose among the other alternatives, assign this PSF level.	

Additional comments (e.g. assumptions)	**No additional comments**

4.3.2. Phase II: The Elicitation Phase

The elicitation phase of the regular determination of the PSF multiplier values is to a large extent similar to the elicitation phase for the validation of PSF multiplier values, see Section 4.2.2.

4.3.3. Phase III: The Calculation Phase

The calculation phase of the regular determination of the PSF multiplier values is to a large extent similar to the calculation phase for the validation of PSF multiplier values, see Section 4.2.3.

5. DISCUSSIONS

There are a range of issues that are open for debate related to the development of the user guidelines for expert judgments. However, the discussion will be limited to the following issues:

1. Selection of experts for validation/evaluation of NHEPs
2. Individual elicitation sessions followed by a group session in estimating NHEPs
3. Selection of experts for regular determination of PSF multiplier values

5.1. Selection of experts for validation/evaluation of NHEPs

We have indicated, in Section 4.1.1, that the experts could be operators, supervisors, instructors on simulators, or HRA analysts. But the question is; if we cannot come up with objective data that supports the NHEP estimates (for diagnosis and action), who is actually the expert having knowledge about the probability of human error for "a typical" diagnosis or action task within the petroleum industry under nominal conditions (i.e. nominal values of the PSFs)? Is it possible to find experts with sufficient knowledge to justify a validation of the existing SPAR-H NHEP values, or which justifies a change in the existing SPAR-H NHEP values?

At the moment we cannot provide an answer to these questions. The approach we will apply is 1) to try to come up with objective data, 2) to fully understand the background for the values used within the nuclear industry, 3) to thoroughly consider who the appropriate experts might be.

5.2. Individual elicitation sessions followed by a group session in estimating NHEPs

In Section 4.1.2, we stated that the elicitation will be carried out individually in an elicitation session using the questionnaire, and alternatively followed by a group session at the end. It is in many cases seen as a possible improvement of the estimates, if the experts are given the opportunity to explain their judgments, including their arguments, to the other experts, followed by a possibility to adjust their judgments based on the discussions. It can still be individual judgments, which is then combined in the calculation phase, or it can be reached consensus between the experts during the group session.

The challenges with a group session are that one single expert may dominate the discussions too much, and even the facilitator/analyst may influence the discussions in an inappropriate manner. The analyst/facilitator must remain neutral and not actively participate in the judgments to avoid that the estimates are affected [22].

If both an individual session and a group session have been carried out, and even a consensus has been reached; which of the three estimates (combined individual estimate, combined adjusted individual estimate, and consensus estimate) should be used in the further calculations? We suggest that this is up to the analyst to decide, but s/he also need to justify the choice, e.g. to stick to the original individual estimates due to an inappropriate group session.

5.3. Selection of experts for regular determination of PSF multiplier values

For the regular determination of PSF multiplier values, the experts to use depend on the project phase, i.e. whether it is an installation in the operations phase or in a "pre-operations"/early phase (design or construction phase). We stated, in Section 4.3.1, that the most suitable expert would probably be an operator carrying out exactly the task in question on an installation in the operation phase, or a supervisor (senior operator) of such operators.

If the judgments are performed during the design phase, then it will not be possible to achieve the same level of plant-specific analysis. Some operators will normally already have been assigned to positions on the installation under design/construction, and can act as experts; however, they will mainly base their judgments on other installations from which they have experience, and not necessarily on this new installation. Factors such as "procedures" and "ergonomics/HMI" may not have been decided on in the project.

It is quite usual that assigned operational personnel are used to provide input to various analyses during design and construction. This is what they do in addition to training, while they are waiting for the new installation to become operational, which means that they are "available and motivated" to provide expert judgments.

Judgments made in early project phases should be updated as soon as new information are obtained, e.g. on "missing" PSFs (such as "procedures" and "ergonomics/HMI").

6. CONCLUSION

Three user guidelines for expert judgments related to validation and evaluation of NHEPs, validation and evaluation of PSF multiplier values (and levels), and regular determination of PSF multiplier values have been prepared.

Even with limited resources we aim for systematic and transparent expert judgment processes. Especially for the determination of PSF multiplier values in regular use of the method, the expert judgments need to be cost-effective, since they are a minor part of the total HRA, which in turn is a minor part of a QRA.

There is quite a lot of flexibility and freedom in designing an expert judgment process; thus, there is no single best way to do this. The processes and user guidelines have been developed based on an existing generic guideline for expert judgments, experience in developing simplified expert judgment procedures, and certain constraints in the specific cases and the Petro-HRA project.

One important part of the design of the user guidelines for regular determination of PSF multiplier values is the extension of current SPAR-H analyst worksheets to two sets of overlapping worksheets; expert judgment worksheets and the analyst worksheets, the first representing the expert judgment questionnaire.

Acknowledgements

The present study was supported by the Petromaks Programme of the Research Council of Norway, Grant no. 220824/E30 “Analysis of human actions as barriers in major accidents in the petroleum industry, applicability of human reliability analysis methods (Petro-HRA)”. Financial and other support from The Research Council of Norway, Statoil ASA and DNV are gratefully acknowledged. This paper represents the opinion of the authors, and does not necessarily reflect any position or policy of the above mentioned organizations.

References

[1] A. D. Swain. *"Human Reliability Analysis: Needs, Status, Trends and Limitations"*, Reliability Engineering and System Safety, 29, 1-4, pp. 301-314, (1990).
[2] *"Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners"*, Office of Safety and Mission Assurance, NASA Headquarters, Washington, DC, August 2002.
[3] K. Øien, N. Paltrinieri and I. Wærø. *"Studies for the quantification by an expert group – familiarization and problem description"*, Petro-HRA report, Norwegian Research Council Project 220824/E30, 2013-06-20. Restricted.
[4] The Petro-HRA project: http://www.sintef.no/SINTEF-Teknologi-og-samfunn/Prosjekter-i-SINTEF-Teknologi-og-samfunn/2012/Petro-HRA/
[5] D. Gertman, H. Blackman, J. Marble, J. Byers, C. Smith, *"The SPAR-H Method"*, NUREG/CR-6883, Washington, DC: US Nuclear Regulatory Commission, 2005.

[6] J. Bell, J. Holroyd, *"Review of human reliability assessment methods"*, Report No. RR679, Health and Safety Laboratory, HSE Books, Sudbury, UK, (2009).
[7] D. E. Embrey, P. C. Humphreys, E. A. Rosa, B. Kirwan, K. Rea, *"SLIM-MAUD: An approach to assessing human error probabilities using structured expert judgement"*, Report No. NUREG/CR-3518. 1984, U.S. Nuclear Regulatory Commission, Washington DC, (1984).
[8] P. Humphreys, *"Human Reliability Assessor's Guide"*, Human Factors in Reliability Group, AEA Technology, London, (1995).
[9] *"Technical basis and implementation guidelines for A Technique for Human Event Analysis (ATHEANA)"*, Report No. NUREG-1624, U.S. Nuclear Regulatory Commission, Washington DC, (2000).
[10] J. Kim, W. Jung, Y. S. Son, *"The MDTA-based method for assessing diagnosis failures and their risk impacts in nuclear power plants"*, Reliability Engineering & System Safety, 93(2), 337-349, (2007).
[11] B. Reer, V. N. Dang, S. Hirschberg, "*The CESA method and its application in a plant-specific pilot study on errors of commission"*, Reliability Engineering & System Safety, 83(2), 187-205, (2004).
[12] H. Pesme, P. Le Bot, P. Meyer, *"Little stories to explain human reliability assessment: a practical approach of the MERMOS method"*, in: IEEE eighth conference on human factors and power plants, 26-31 August 2007, Monterey, California, (2007).
[13] A. D. Swain, H. E. Guttmann, *"Handbook of human reliability analysis with emphasis on nuclear power plant applications"*, Report No. NUREG/CR-1278, U.S. Nuclear Regulatory Commission, Washington DC, (1983).
[14] J. C. Williams, *"HEART-a proposed method for assessing and reducing human error"*, in: Proceedings of a Symposium on the Achievement of Reliability in Operating Plant, Safety and Reliability Society. NEC, Birmingham, (1985).
[15] Spurgin A. J., *"Human Reliability Assessment, Theory and Practice"*, Boca Raton, Florida, (2010).
[16] A. M. Whaley, D. L. Kelly, R. L. Boring, W. J. Galyean, *"SPAR-H Step-by-step Guidance"*, Idaho National Laboratory, Risk, Reliability, and NRC Programs Department, 2011.
[17] M. Musharraf, J. Hassan, F. Khan, B. Veitch, S. MacKinnon, S. Imtiaz, *"Human reliability assessment during offshore emergency conditions"*, Safety Science, 59, 19-27, (2013).
[18] L. Podofillini, V. N. Dang, *"A Bayesian approach to treat expert-elicited probabilities in human reliability analysis model construction"*, Reliability Engineering and System Safety, 117, 52-64, (2013).
[19] R. L. Boring. *"How Many Performance Shaping Factors are Necessary for Human Reliability Analysis?"*, 10th International Probabilistic Safety Assessment & Management – PSAM10, 7-11 June 2010, Washington USA.
[20] R. Boring, D. Gertman, J. Joe, J. Marble, W. Galyean, L. Blackwood, H. Blackman, *"Simplified Expert Elicitation Guideline For Risk Assessment Of Operating Events"*, Report No. INL/EXT-05-00433, Idaho National Laboratory, Idaho Falls, Idaho, (2005).
[21] K. Øien, O. J. Klakegg, P. Hokstad, R. Rosness, *"Handbook for performing expert judgments"*, Report No NTNU 96004, Norwegian University of Science and Technology, 1996 (in Norwegian).
[22] K. Øien, P. R. Hokstad, *"Handbook for performing expert judgements"*, SINTEF Industrial Management: SINTEF Report STF38 A98419, Trondheim, Norway, 1998.
[23] J. B. Soll, J. Klayman, *"Overconfidence in interval estimates"*, Journal of experimental psychology: learning, memory and cognition, 30, 299-314, (2004).