Usage of Human Reliability Quantification Methods

Miroljub Grozdanovic

Faculty of Occupational Safety, University of Nis, Serbia and Montenegro

Human reliability quantification (HRQ) methods are becoming increasingly important in risk and accident assessment in systems these terms are usually related to (hi-tech industrial systems, including nuclear and chemical plants).

These methods began to intensively develop after numerous accidents caused by human error or inadequate activity of people who controlled and managed complex technological processes. For already existing systems, but also for new ones, it is important to assess the possibility of an accident. Determination of possible preventive activities, which include the influence of human error on the safety of a system, is also required. These are the main goals of the HRQ method.

Using Absolute Probability Judgment (APJ) and Success Likelihood Index Methods (SLIM) HRQ techniques in control and management centers in electro-power systems in Belgrade and railway traffic in Nis (both in Serbia and Montenegro) are shown in this paper.

human error   expert method   operator's reliability   control center

1. INTRODUCTION

Human reliability quantification (HRQ) techniques all quantify human error probability (*HEP*), which is a measure of human reliability assessment. Industrial studies of performance and accidents would be the ideal source of human error data. Other sources are simulation data and data derived from literature on human performance.

The term human error has been pragmatically defined by Swain (1989) as follows: "any member of a set of human action or activities that exceeds some limit of acceptability, i.e. an out of tolerance action where the limits of performance are defined by the system" (as cited in [1]). The effects of human error on system performance have been demonstrated most vividly by large-scale accidents. Since the intention here is merely to highlight the human-error aspects of accidents, these brief descriptions, showed in Table 1 are, for our purposes, appropriate [1]. For other descriptive references to a range of accidents see Reason [2].

TABLE 1. Accidental Events

Year	Event	Domain
1966	Aberfan disaster	Mining
1972	Crash of the BEA Trident 1	Aviation
1973	Paris air disaster	Aviation
1974	Flixborough disaster	Chemical
1975	Browns Ferry fire	Nuclear power
1975	Dutch States Mines explosion	Chemical
1976	Seveso incident	Chemical
1977	Ekofisk Bravo blowont	Offshore
1978	Bantry Bay disaster	Petrochemical
1979	Three Mile Island accident	Nuclear power
1984	Bhopal catastrophe	Chemical
1985	Davis Besse incident	Nuclear power
1986	Challenger Space Shuttle disaster	Space
1986	Chernobyl	Nuclear power
1989	Exxon Valdes accident	Oil
1991	Hevsi accident	Oil
1996	Sea Empress accident	Oil
2003	Secuan explosion	Petrochemical
2004	Tianjuan explosion	Chemical
2004	Sinuuidjua accident	Railway trafic

Correspondence and requests for offprints should be sent to Miroljub Grozdanovic, Faculty of Occupational Safety, University of Nis, Serbia and Montenegro. E-mail: <mirko@znrfak.znrfak.ni.ac.yu>.

Current accident experience suggests that so-called high-risk industries are still not particularly well protected from human error. This, in turn, suggests the need both for the means of properly assessing risk attributable to human error and for ways of reducing system vulnerability to human error impact. These are the primary goals of Human Reliability Assessment (HRA) achieved by its three principal functions of identifying what errors can occur, deciding how likely they are to occur and enhancing human reliability by reducing the likelihood of those errors. Human reliability assessment clearly has an important role to play, and this role is likely to extend to many industries, wherever human errors can propagate within systems, to lead to unacceptable events. HRA is a hybrid area, arising out of the disciplines of engineering and reliability on the one hand, and psychology and ergonomics on the other. The former require human error probabilities to fit neatly into the logical mathematical framework of probabilistic safety analysis (PSA), and the latter urge more detailed and theoretically valid modeling of the complexity of the human operator. Whilst a good deal of research has been carried out in the field of identifying human error and particularly classifying errors, few practical techniques have been developed in Yugoslavia for use in risk assessments [3, 4, 5]. It is likely that future research and development will focus on the development of such techniques.

Human error analysis is arguably the most critical part of human reliability analysis since if a significant error is omitted at this stage then it will not appear subsequently in the analysis and hence the results may seriously underestimate the effects of human error on the system.

Identification of performance shaping factors (PSF) which affect performance can obviously be usefully considered during the human error identification phase although frequently they are not identified until the quantification phase.

Therefore, there is a data problem. Such difficulties have led to the development of non-data-dependent approaches, namely to the use of expert opinion. This is by no means necessarily a bad thing, and expert opinion has been used successfully in other areas, and is in any case used at least occasionally in probabilistic safety assessment where similar problems often exist.

Where personnel do not have direct experience of the event in question, if the event is similar to one for which an expert does have knowledge and experience, such an expert, or group of experts, may be able to construct a reasonably accurate estimate of the likelihood of the occurrence of such an event.

In both these cases, a critical assumption is being made that the expert has useful and accurate knowledge of the problem domain being investigated. This is known as having substantive expertise.

The adage of "two heads are better than one" is nowhere more true than in Absolute Probability Judgment (APJ), and in light of this, single-expert approaches are not considered further.

There are several ways of aggregating several experts' opinions: they can estimate alone, with their opinions then aggregated mathematically; or they can estimate alone but have limited discussions for clarification purposes; or they can meet as a group and discuss their estimates until they reach a consensus.

The four APJ group approaches are described briefly here.

- Aggregated individual method; this method entails that the experts do not meet but make estimates individually. These estimates are then aggregated statistically by taking the geometric mean of all the individual estimates for each task.
- Delphi method; experts make their assessments individually and then all the assessments are shown to all the experts.
- Nominal group technique; this method is similar to the delphi method, but after the group discussion, each expert makes his or her own assessment. These assessments are then statistically aggregated.
- Consensus-group method; in this method, each member contributes to the discussion, but the group as a whole must then arrive at an estimate upon which all members of the group agree.

In practice, it will be up to the practitioner carrying out the study to decide which method to

use by assessing the requirement for information sharing and accuracy in the estimates made against the possible practical difficulties involved in co-locating and "managing" a group of experts.

In a review by Kirwan et al., eight HRQ techniques were qualitatively assessed [6]. These were: APJ, Paired Comparisons (PS), Teseo, Technique for Human Error Rate Prediction (THERP), Human Error Assessment and Reduction Technique (HEART), Influence Diagrams Approach (IDA), Human Cognitive Reliability Model (HCR) and Success Likelihood Index Method (SLIM).

Four of the techniques (APJ, PC, IDA and SLIM) use a group of expert judges to evaluate *HEP*.

Within the scope of this paper it is not possible to review all techniques, but two are reviewed, namely, SLIM and APJ.

2. METHODOLOGY

The APJ approach is conceptually a most straight forward HRQ approach. It simply assumes that people can either remember or, better still, estimate directly the likelihood of an event—in this case, a human error. When it comes to risk

assessments for existing plants or systems, it is arguable that the more experienced personnel will have a reasonable memory of their own errors, as well as of other operators' errors and their rates of occurrence. Steps of APJ procedure are shown in Figure 1.

These steps are now detailed here.

- Step 1. The number of experts needed to make the required judgments cannot be stated unequivocally. As many experts as practicable should try to participate. Many authors suggest six experts would be sufficient for a direct estimation, although more would be preferable. In practice, however, financial and other constraints often lead to the use of a smaller group of only three or four experts. In general, if a group consensus is aimed for, a group of 4–6 people is preferable, since problems are likely to occur with larger groups.
- Step 2. Well-defined task statements are a critical aspect of the APJ-estimation procedure. The more fully the tasks are specified, the less they will be open to individual interpretation by the experts when they are making their judgments. The levels of detailed definitions will vary according both to the nature of the task itself and to the final use of the *HEP* estimate.
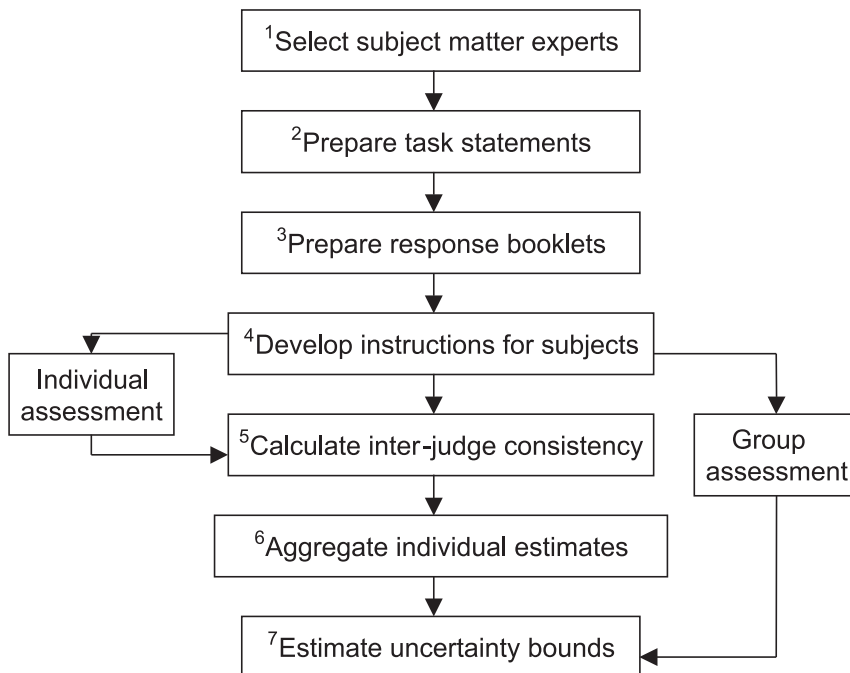

Figure 1. Steps in the Absolute Probability Judgment (APJ) procedure.

- Steps 3 and 4. A key consideration when using the APJ approach is the type of scale on which experts will indicate their judgments. It is important that the chosen scale be of sufficient detail to allow the degree of sensitivity of the expert to individual differences to be indicated. The scale values must also reflect the estimated range of the true error probabilities of the tasks, where these are known. If they are not known, then a range of 10^0 to 10^{-6} is sufficient.
- Step 5. Experts are asked either to work through their booklets or, when operating in a group-consensus mode, to discuss each task in turn and arrive at a consensus estimate. It may be useful, when the consensus mode is in operation, to let the experts review all the tasks and start on one they feel confident that they can assess. Individual *HEP* estimates should only be used if there is a reasonable level of agreement between the experts. To make subsequent calculation easier, the set of *HEP*s obtained from the experts are then translated into their logarithmic equivalents.
- Step 6. If a consensus group is not used, and if the level of agreement between judges is adequate, it will next be necessary to aggregate the different individuals' estimates for each *HEP*. This is achieved by taking the geometric mean of the individual estimates.
- Step 7. Uncertainty bounds are calculated using the following formulae.

Upper and lower uncertainly bounds are equivalent to:

$$\log HEP + 2\,SE$$

where *SE* is standard error $\left(\sqrt{\dfrac{V(\log HEP_i)}{m}}\right)$,

$$V(\log HEP_i) = \frac{\left[m\sum_{j=1}^{m}(\log HEP_{ij})^2\right] - \left[\sum_{j=1}^{m}(\log HEP_{ij})^2\right]}{m(m-1)},$$

where *m*—number of experts, *n*—number of events.

SLIM also uses expert judges but the judges are asked to consider what factors affect performance, and from the assessment of those factors and the modeling of their influence on performance, they then determine the *HEP*. They are assisted by the analyst in creating a quantitative causal model of the influence of these factors on the *HEP*. Typical performance shaping factors (PSFs) utilized are motivations, the quality of the interface design, the degree of training and adequacy of procedures.

SLIM can best be explained using the example of human reliability assessment in the case of operator activities in the control and management center of the Yugoslav railway.

The expert panel would typically consist of four operators in the control and management center of the Yugoslav railway in Nis (with 10 years of experience), one human factors analyst and a reliability analyst familiar with the system, who also has some operational experience.

The panel is initially asked to identify a set of PSFs, which are any factors relating to the individuals, environment, or task, which affect performance positively or negatively. The expert panel could be asked to nominate the most important or significant PSFs for the scenario under investigation. In this example it is assumed the panel identifies the following major PSFs as affecting human performance in this situation: training, design of display boards, and design of control panels, procedures, motivations and illumination. The panel is then asked to consider other human errors possible in this scenario and for each one to decide to what extent each PSF is optimal or sub-optimal for that task in the situation being assessed. The rating of whether a task is optimal or sub-optimal for a particular PSF is made on a scale of 1 to 9, with 9 being *optimal*.

SLIM and the decision analysis technique are based upon the multi-attribute rating technique, which simply proposes that preference can be derived as a function of the sum of the weightings multiplied by their rating for each item (human error). SLIM does this and calls the resultant preference index a success likelihood index (SLI). However the SLIs are not yet probabilities. Rather they are indications of the relative likelihood of the different errors. Thus SLIs show the ordering of likelihood of the different errors, but do not yet define the absolute probability values. In order to

transform the SLIs into *HEP*s, it is necessary to "calibrate" the SLI values.

3. RESULTS

Using the APJ method, we assessed an operator's error in the control and management center in electro-power systems in Belgrade, Serbia and Montenegro. An assessment of eight cases of error ($n = 8$) was made by four experts ($m = 4$). An example set of human-error probabilities (*HEP*s) is shown in Table 2. To make subsequent calculation easier, the set of *HEP*s obtained from the expert were then translated into their logarithmic equivalents, resulting in the set of figures in Table 3.

The computational instruction for an analysis of variance is set out here:

1. Calculate the column totals (t): –9.09, –7.95, etc.
2. Calculate the row totals (r): –19.04, –12.48, etc.
3. Calculate the grand total (T): –71.55
4. Calculate the correction term (C): $C = T^2/mn$, therefore $C = (71.55)^2/32 = 159.98$
5. Calculate the sum of the squares (x^2) of the raw scores: $x^2 = 180.67$
6. Calculate the total sum of the squares (*TSS*): $TSS = 20.69$
7. Calculate the between column sum of squares (t^2): $t^2 = -78.02$
8. Calculate the between row sum of squares (r^2): $r^2 = 175.22$
9. Calculate the residual sum of squares: $SS = -76.51$
10. Enter the appropriate degrees of freedom into the summary table:
    Columns differential   7
    Rows differential   3
    Total differential   31
    Residual differential   21
11. Calculate the variance estimates by dividing each of the sums of squares by the appropriate degrees of freedom:
    Column (event) variance   –11.15
    Row (expert) variance   58.40
    Residual variance   –3.64
12. Calculate the F ratios
    F (columns)   3.06
    F (rows)   –16.04
13. The last step is to determine the intra-class correlation coefficient, according to the following formula:

$$r = \frac{F - 1}{F + (n - 1)} \quad \Rightarrow \quad r_1 = 0.20 \quad r_2 = -2.16,$$

where F is the ratio for the events factors.

Using SLIM we assessed an operator's errors in the control and management center of the Yugoslav railway. The ratings obtained for five human errors under analysis are as shown in Table 4 [7].

TABLE 2. Absolute Probability Judgment (APJ)—Derived Human Error Probabilities

Expert ($m = 4$)	Event ($n = 8$)							
	$p_1(L_G)$	$p_1(L_D)$	$p_1(D_G)$	$p_1(D_D)$	$p_2(L_G)$	$p_2(L_D)$	$p_2(D_G)$	$p_2(D_D)$
1	0.00059	0.00830	0.00560	0.01000	0.00056	0.00750	0.00880	0.00990
2	0.05200	0.06000	0.00590	0.06800	0.00600	0.08100	0.10000	0.00600
3	0.05000	0.00400	0.00700	0.00600	0.00056	0.07000	0.00060	0.10000
4	0.00055	0.00590	0.00065	0.00700	0.00059	0.00080	0.00085	0.00095

TABLE 3. Log Human Error Probabilities

Expert ($m = 4$)	Event ($n = 8$)								Total
	1	2	3	4	5	6	7	8	
1	–3.23	–2.09	–2.26	–2.00	–3.26	–2.13	–2.06	–2.01	–19.04
2	–1.29	–1.23	–2.23	–1.17	–2.23	–1.10	–1.00	–2.23	–12.48
3	–1.31	–2.40	–2.16	–2.23	–3.26	–1.16	–3.23	–1.00	–16.75
4	–3.26	–2.23	–3.19	–2.16	–3.23	–3.10	–3.08	–3.03	–23.28
Total	–9.09	–7.95	–9.84	–7.56	–11.98	–7.49	–9.37	–8.27	–71.55
Average	–2.27	–1.99	–2.46	–1.89	–2.99	–1.87	–2.34	–2.07	

TABLE 4. Ratings Obtained for Five Human Errors

Errors	Performance Shaping Factors (R)					
	Training	Design of Display Boards	Design of Control Panels	Procedures	Motivation	Illumination
1 Inadequate notiception of sign backgrounds colors	7	9	3	3	4	6
2 Inadequate notiception signs because of its dimension	4	8	6	2	3	8
3 Alarm ignored	8	6	6	9	5	3
4 Missed information because of inadequate viewing angles	7	9	5	3	6	8
5 Inadequate managing action	7	6	9	8	5	4
Σ	33	36	29	25	23	29

Weightings for the PSFs can be obtained directly from these considered opinions, as follows, normalized to sum to unity (Table 5).

TABLE 5. Normalized Value

Performance Shaping Factors	Weighting (W)
Design of display boards	0.30
Training	0.25
Design of control panels	0.15
Illumination	0.15
Procedures	0.10
Motivation	0.05
Σ	1.00

SLI calculations are showed in Table 6.

have derived such a calibration relationship, suggesting a logarithmic relationship of the form:

$$\log(HEP) = a\,SLI + b.$$

If two more tasks which had *HEP*s of 0.5 and 10^{-4} respectively were assessed and were given SLIs of 4.00 and 6.00 respectively, then the derived equation would be:

$$\log(HEP) = -1.85\,SLI + 7.1.$$

The *HEP*s would then be:

$$\log(HEP1) = -4.56 \Rightarrow HEP1 = 0.0000270$$
$$\log(HEP2) = -3.72 \Rightarrow HEP2 = 0.0001900$$
$$\log(HEP3) = -4.56 \Rightarrow HEP3 = 0.0000270$$
$$\log(HEP4) = -5.85 \Rightarrow HEP4 = 0.0000014$$
$$\log(HEP5) = -5.02 \Rightarrow HEP5 = 0.0000095$$

TABLE 6. SLI Calculation

Weighting (W)	PSF (R)	SLI = W·R	1	2	3	4	5
0.30	Design of display boards	0.30 (9, 8, 6, 9, 6)	2.70	2.40	1.80	2.70	1.80
0.25	Training	0.25 (7, 4, 8, 7, 7)	1.75	1.00	2.00	1.75	1.75
0.15	Design of control panels	0.15 (3, 6, 6, 5, 9)	0.45	0.90	0.90	0.75	1.35
0.15	Illumination	0.15 (6, 8, 3, 8, 4)	0.90	1.20	0.45	1.20	0.60
0.10	Procedures	0.10 (3, 2, 9, 3, 8)	0.30	0.20	0.90	0.30	0.80
0.05	Motivation	0.05 (4, 3, 5, 6, 5)	0.20	0.15	0.25	0.30	0.25
		Σ	6.30	5.85	6.30	7.00	6.55

Notes. SLI—success likelihood index, PSF—performance shaping factors.

In this case, the lowest SLI is 5.85, suggesting that inadequate notiception because of its dimension is still the most likely error.

In order to transform the SLIs into *HEP*s, it is necessary to calibrate the SLI values (the paired comparison technique also requires this calibration using the same basic formula). Two earlier studies

4. CONCLUSION

The APJ method is relatively quick to use, and yet it also allows as much detailed discussion as the experts think fit; this kind of discussion, if documented, can often itself be qualitatively useful.

Discussions can also be turned towards a consideration of how to achieve error reductions. In such a situation, the group becomes like a HAZOP (Hazard and Operability Study) group, and can develop some highly credible and informed suggestions for improvements. This development is also beneficial where the group members are themselves operational staff, since this fact would improve the chances of such recommendations being accepted and then properly implemented. The APJ method is prone to certain biases, as well as to personality group problems and conflicts, which, if not effectively countered (e.g., by a facilitator), can significantly undermine the validity of the technique.

All forms of APJ whether group or individual are prone to particular biases which can detract from the accuracy of the experts' assessments. One of the major ways of reducing the problem of biases in expert judgment is to employ a facilitator during the experts' group session. The primary function of the facilitator is to try to overcome these biases. A secondary function, therefore, of the facilitator is to overcome any personality conflicts, or other problems, which may occur in small groups, and influence the assessment process.

The following only deals with some of the more well-known biases, and not the various personality conflicts which can occur:

- The overconfidence bias; this causes uncertainty bands to be too narrow; i.e., it generally causes underestimation of very high failure probabilities, on the one hand, and overestimation of very low failure probabilities on the other.
- The availability bias; here, the value of the estimate made by the experts involved is affected by the ease with which they can bring to mind previous occurrences of the events in question.
- The anchoring bias; an expert, or expert group, starts with some initial value suggested by one member, and adjusts this value so as to derive the best estimate, frequently failing, however, to adjust it to a sufficient extent.
- The motivational bias; this occurs when an expert, or a group of experts, have a vested interest in obtaining probabilities of a certain value or, e.g., low probabilities versus high ones.

In this paper the rationale underlying SLIM is applied. The computerized version, is known as SLIM-MAUD (SLIM using Multi Attribute Utility Decomposition). Due to the mathematics in the software, which is present partly to avoid such bias, it will produce slightly different values (*HEP*s) than the hand calculated method used earlier. In particular the simple summary of weightings and ratings is refined in several ways according to the multi-attribute utility theory.

References

1. Kirwan B. A guide to practical human reliability assessment. London, UK: Taylor & Francis; 1994.

2. Reason JT. Human error. Cambridge, UK: Cambridge University Press; 1990.

3. Grozdanovic M. Human factors and preventive engineering. Facta Universitatis (Series: Working and Living Environmental Protection) 2001;2(1):39–50.

4. Grozdanovic M, Savic S. Methodology for human reliability quantification, In: Proceedings of the XVth Triennial Congress of the International Ergonomics Association and the 7th Joint Conference of Ergonomics Society of Korea/Japan Ergonomics Society. Seoul, Korea: Ergonomics Society of Korea; 2003. vol. 4, p. 5–7.

5. Savic S, Grozdanovic M. Risk management of operators workplace in control rooms in the coal mines. In: Proceedings of the XVth Triennial Congress of the International Ergonomics Association and the 7th Joint Conference of Ergonomics Society of Korea/Japan Ergonomics Society. Seoul, Korea: Ergonomics Society of Korea; 2003. vol. 5, p. 435–8.

6. Kirwan B, Embrey, DE, Rea K. The human reliability assessor's guide (Report RTS 88/95Q). Culcheth, Cheshire, UK: National Centre for Systems Reliability (NCSR), United Kingdom Atomic Energy Authority (UKAEA); 1988.

7. Grozdanovic M, Savic S. Assessment of human error probability using success likelihood index method. In: 5th International Conference on Reliability, Maintainability and Safety, Dalian, China; 2001.