Reliability Engineering and System Safety 230 (2023) 108903

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

A traceable process to develop Bayesian networks from scarce data and expert judgment: A human reliability analysis application

Luca Podofillini [a,*], Bernhard Reer [b], Vinh N. Dang [a]

[a] Laboratory for Energy Systems Analysis, Paul Scherrer Institute, Risk and Human Reliability Group, Forschungsstrasse 111 PSI, Villigen, Switzerland
[b] Swiss Federal Nuclear Safety Inspectorate ENSI, Eidgenössisches Nuklearsicherheitsinspektorat, Sektion Probabilistische Sicherheitsanalysen and Accident Management PSAM, Industriestrasse 19 AG, Brugg, Switzerland

ARTICLE INFO

Keywords:
Bayesian belief networks
Human reliability analysis
Event and accident analysis
Errors of commission
Expert judgement

ABSTRACT

The present paper develops a Bayesian Belief Network (BBN) for quantification of aggravating actions, as outcomes of inappropriate decisions, to be integrated in probabilistic safety assessment (PSA) models (i.e., the so-called errors of commission, EOCs). The BBN connects analyst ratings on influencing factors to the error forcing impact of a specific scenario, supporting the CESA-Q method (the Quantification module of the Commission Error Search and Assessment method). While contributing to the quantification of EOCs, this paper presents a novel process for the quantification of the BBN parameters (the Conditional Probability Distributions, CPDs), striving for traceable integration of expert knowledge and (scarce) data, in the form of retrospective analyses of operational events involving EOCs. The process combines the functional interpolation method for populating CPDs and Bayesian updates to adjust the BBN response to the available evidence. A first, prior BBN is developed, then sequentially updated to adjust to two data sets. This allows some intermediate validation and puts forwards the steps for future BBN updates as new EOC events (or new analyst assessments) become available.

1. Introduction

Integration of the human element in risk analysis is required for a realistic and informative risk profile when technical systems involve human-machine interactions. In this respect, Human Reliability Analysis (HRA) is the discipline to address safety-relevant interactions, potential failures and the factors driving performance [1,2]. An important output of HRA is the quantification of human failure probabilities, then incorporated in the overall risk models, e.g., fault trees and event trees of Probabilistic Safety Assessment (PSA).

State-of-the-art HRA addresses failures to perform actions required by procedures and training, for example in coping with disturbances and accidents [3,4]. Traditionally, these failures are referred to as "Errors of Omissions" (EOOs), emphasizing the non-performance of the required actions. An important challenge to a more comprehensive risk profile is the incorporation of inappropriate actions, not required in the specific scenarios, that erroneously or unintentionally, aggravate the state of the system (e.g., a nuclear power plant, a chemical plant, an aircraft): examples include terminating running injection pumps, inhibiting automatic initiation signals [4–6]. In the HRA and PSA terminology, these actions are often referred to as "Errors of Commission" (EOCs).

Experience has accumulated over the years on how to prioritize the identification of inappropriate actions to those that are plausible, motivated, and risk-important [6–8]. Of particular concern are the EOCs resulting from inappropriate decision-making (mistakes, errors of intention, in Reason's taxonomy [9]); the inappropriate understanding at their origin tends to affect the interpretation of the system's response, making these decisions more difficult to reconsider and recover [10]. The quantification issue concerning inappropriate actions due to slips or lapses (e.g., pushing wrong button in proximity of the correct one) is considered of different nature. For these errors, data-based probabilities are available (e.g., [11]). In addition, modern system designs have ergonomic provisions for the prevention of inadvertent activations (e.g., activation requires pressing two buttons simultaneously).

Quantification of probabilities of decision-making failures represents an open issue because these decisions are triggered by very specific combinations of contextual influences and plant conditions, e.g., misleading indications or procedural instructions, training biases, unexpected plant conditions or scenario evolutions, conflicting goals [10,12,13]. Instead, traditional HRA methods [1,2] address contextual influences one-by-one, via sets of Performance Shaping Factors (PSFs, e.g., procedural guidance, training, human-machine interface). Correspondingly, the underlying mathematical models for the quantification of

* Corresponding author.
E-mail address: luca.podofillini@psi.ch (L. Podofillini).

https://doi.org/10.1016/j.ress.2022.108903
Received 11 August 2022; Received in revised form 11 October 2022; Accepted 14 October 2022
Available online 17 October 2022
0951-8320/

Nomenclature

AD	adverse distraction (CESA-Q Situational feature)
AE	adverse exception (CESA-Q Situational feature)
BBN	Bayesian belief network
BP	benefit Prospect (CESA-Q adjustment factor)
CESA-Q	commission Error Search and Assessment method - Quantification module
CPD	conditional probability distribution
DP	damage potential (CESA-Q adjustment factor)
EFI	error forcing impact
EOO	errors of omissions
EOC	errors of commission
HRA	human reliability analysis
MI	misleading indication or instruction (CESA-Q situational feature)
PR	personal redundancy (CESA-Q adjustment factor)
PSA	probabilistic safety assessment
PSF	performance shaping factor
TP	time pressure (CESA-Q adjustment factor)
VD	verification difficulty (CESA-Q adjustment factor)
VE	verification effort (CESA-Q adjustment factor)
VH	verification hints (CESA-Q adjustment factor)
VM	verification means (CESA-Q adjustment factor)

failure probabilities is often multiplicative: each PSF contributes independently of the other PSFs, with a multiplier to a failure probability in nominal performance conditions, e.g., [11,14–16].

To address the modeling shortcoming, including but not limited to the above, Bayesian Belief Networks (BBNs) [17–19] are becoming increasingly popular in HRA research [20,21]. BBNs are probabilistic graphical models to represent and quantify relationships among different types of interrelated variables. The primary use of BBNs is the representation of knowledge and decision support under uncertainty; their application is established in diverse areas such as medical diagnosis and prognosis, engineering, finance, information technology, natural sciences [18]. BBNs have a number of attractive features for HRA and for risk analysis in general: their intuitive graphical representation, the possibility of combining diverse sources of information (including subjective knowledge), their ability to represent causal effects [19–21]. BBNs allow modeling strong factor effects and interactions, provided that these effects can be quantified: this potentially overcomes the assumption of some methods of independent factor effects [20]. Examples of BBN applications in HRA are abundant in the recent literature: to capture causal relationships among factors influencing human failures [21–23], as well as overall accidents [24,25]; to incorporate organizational aspects and safety culture in risk analysis [26,27], to capture dynamic aspects of human performance [28,29], and different types of dependencies (among influencing factors and among failures) [30,31]. In [32], Bayesian network models are proposed as the foundation of third generation HRA methods, supporting rigorous treatment of causality relationships across influencing factors and aggregation of multiple data types.

While contributing to the quantification of EOCs, the present paper puts forward a novel process for the quantification of the BBN parameters (the Conditional Probability Distributions, CPDs), combining expert knowledge and (scarce) data. Common to several other studies [22–24], the available data is in the form of records of human failures and corresponding influencing factors. Differently from the cited work, the amount of data is not sufficient for meaningful statistical analysis, e.g., correlation analysis as in [22,23], or statistical inference in [24]. Another difference with the typical HRA-BBNs literature is that, in the present work, the influencing factors are associated multiple possible levels and therefore are coded as multi-state nodes (variables) in the BBN model. Instead, in most cases the influencing factors are coded as binary factors (e.g., present / not present, adequate / not adequate). Multi-state nodes increase the number of CPDs to be quantified, further increasing the data requirements. On the other hand, as presented in the next Section 2, multi-level factors allow the representation of graded strengths of error-forcing influence of the context on the personnel decision under analysis.

The solution adopted in the present work involves the use of methods for populating BBN CPDs from partial information [33–35], typically adopted when expert judgment is the primary source of information for the CPDs assessment. The thrust of these methods is to elicit only selected model parameters (e.g., selected CPDs or information about factor importance) and derive the remaining parameters via formulas or algorithms, depending on the method. Their fundamental motivation is to avoid eliciting a large number of probabilities, which could lead to biases and inconsistencies in the assessments (these issues are presented in detail in [36,37]).

In particular, this paper presents a BBN-based model to support the quantification module of the Commission Error Search and Assessment (CESA-Q) method [38], developed by some of the authors to address decision-related EOCs. Given a specific EOC scenario, the BBN is intended to support the assessment of the Error Forcing Impact (EFI) of the context on the decision, on a five-level scale. The BBN takes the analyst ratings on the CESA-Q performance factors (called “adjustment factors”) as inputs. The process devised for building the BBN combines operational event analyses (from two data sets, EOC Set I [39] and EOC Set II [40]), expert judgment and a CPD filling method, the functional interpolation method [35] the latter being particularly suitable for medium-sized BBN models (say, not exceeding 20–30 nodes), with multistate nodes and dependent factor effects [35].

The paper is structured as follows. Section 2, next, gives an overview of the CESA-Q method for EOC quantification [38]. Besides briefly introducing the method and the two EOC event Sets I and II [39,40], the section presents the overall motivation for developing the BBN model. Section 3 presents the devised BBN development process, in particular the functional interpolation method [35] and the sequential CPD updates to Posterior I (based on EOC Set I) and Posterior II (based on EOC Set II). Some fundamental concepts about the BBN modeling framework are given too. Section 4 presents the implementation of the BBN development process, with the main results (all parameters of the developed BBN are made available as supplementary material on the paper web page). Section 5 discusses the main achievements of the paper, with emphasis on the capabilities of the BBN modeling framework, on the relevant features of the development process and on the future steps for the EOC quantification work. Conclusions are given at the end.

2. CESA-Q: a method for quantifying errors of commission

The Commission Error Search and Assessment (CESA) method was developed with the focus on identification and prioritization of EOCs [7]; the CESA-Q module came later, to address EOC quantification [38]. CESA-Q focuses on decision EOCs, i.e. for which the inappropriate action is committed as a result of a motivated decision (the action is intentionally made, and its inappropriateness is not known to the operators).

CESA-Q aims at a holistic, high-level characterization of the situation resulting in the inappropriate decision, from the recognition that decision are strongly influenced by the overall context, as opposed by single performance factors (e.g., [10,12,13]). The characterization is done via two layers of factors [39]:

- Situational factors (e.g., misleading indications), which describe at a high level why operators did consider a specific option (which then may turn out to be inappropriate).

Table 1
Correspondence of error-forcing impact, reliability index i, mean probability of EOC in CESA-Q [38].

Error-Forcing Impact	Reliability index i	Mean Prob(EOC|i)[1]
Extremely high	0	1
Very high	1	2.7e-1
High	2	7.2e-2
Low	3	1.9e-2
Very low	4	5.2e-3
None	5	1.4e-3[2]

[1] The mean probability of an EOC event given a context characterized by reliability index i is a function of i: Prob(EOC|i)= exp(-c•i), with the constant c = 1.315, obtained via a statistical analysis of operational events [38].
[2] EOC probabilities below this value are possible. This value conservatively represents the limiting case if no EFI has been identified.

- Adjustment factors (e.g., availability of backup indications or support from procedural guidance, time pressure), which characterize the strength of the EOC-motivating context.

In the first layer, four situations are distinguished: misleading indications or instructions; an adverse condition (e.g., an earlier fault) that makes an action inappropriate (the action would be otherwise appropriate); a distracting occurrence suggesting the need for an action (which, unknown to the operators, has aggravating impact on the course of a plant event); deviations from the recognized rule associated with a notable benefit and typically no adverse consequences.

CESA-Q's situation-based perspective transfers to the second layer, the adjustment factors (Table A1 of Appendix A). For example, the factors "Verification Hint" (VH) and "Verification Means" (VM) aggregate the effects of the procedures, human-machine interface, and training to support the verification of the decision. Typical HRA methods, addressing non-performance of required actions (EOOs), would address separately the adequacy of each of these factors. CESA-Q addresses their effect combined and with the specific focus on verification of the decision. "Factor Benefit Prospect" (BP) is another example, weighing how much the potentially positive consequences of the inappropriate decision may influence the decision.

The CESA-Q factor framework was developed based on the analysis of twenty-six operational events [39], most of which occurred in the 1990s, with the most recent event in 2000. The eighteen events related to situational features "Misleading Indication or Instruction" (MI), "Adverse Exception" (AE), "Adverse Distraction" (AD) constitute EOC Set I (reported in Table A2 of Appendix A). The rest of the events in [39] relate to situational feature "Risky Incentive" and are not addressed in the present paper because of their substantially different type of decision drivers (a different BBN model may be required to capture these events). As reported in Table A2, each entry includes ratings of the eight CESA-Q adjustment factors in the specific event situations as well as an evaluation of the corresponding error-forcing impact on the five-level scale, see column EFI in Table A2 (note the gray columns in Table A2 have been added to support the BBN model and are not part of the [39] analysis, see Section 3.1). The CESA-Q framework has been recently applied to operational events that occurred in the period 2000–2016 [40]: these analyses are reported in Table A3 and constitute EOC Set II (except for the "Risky Incentive" situations). It is important to note that the two EOC sets have been populated by two different analysts (one for EOC Set I and one for EOC Set II) and with large time separation (over 10 years between the two sets), therefore some analyst-to-analyst variability may exist.

The concept for EOC quantification in CESA-Q reflects its situation-based perspective, and does not adopt the PSF-based multiplicative model. As shown in Tables A2 and A3, CESA-Q quantification results in the strength of the Error-Forcing Impact (EFI) of the context under analysis (on a discrete scale, see Table 1), on the basis of the eight adjustment factors. The strength of the impact represents the overall belief regarding the positive or negative effects on the EOC probability. The probability of committing the error is related to the EFI via the so-called reliability index and a functional relationship (Table 1). The applicable EFI is determined by the analyst based on an overall context evaluation, through the adjustment factors. In its original form [38], this is based on a match-and-adjust approach: it involves comparing the EOC under analysis with entries from the CESA database (the 26 events of [39]). The closest entry in the database in terms of pattern of adjustment factor ratings provides the reference probability value for the new analysis. Given the limited number of entries in the database, the identification of a close match is indeed rare. The present work is motivated by the need to support the EFI assessment by an explicit model, to reduce the subjectivity of the match-and-adjust approach. The applicable EFI, and therefore the error probability via Table 1, directly follows through the model from the adjustment factor evaluations (which become the model inputs), without need for additional judgments by the analyst. The need to combine scarce data with expert judgment, the possible dependence in the adjustment factor effects and the factor representation on multi-valued scale led to the choice of Bayesian Belief Networks as the modeling framework.

3. The proposed BBN development process

After a short presentation of the BBN modeling framework, this section gives an overview of the BBN development process devised for this paper (Fig. 1). The next subsections provide more details on the main elements of the process: in Section 3.1, the structure of the BBN and of the evidence (EOC Sets I and II); in Section 3.2, the functional interpolation method [35]; in Section 3.3, the sequential update to BBN Posterior I and II.

A BBN is a probabilistic graphical model whose structure consists in nodes linked by directed arcs [17,18]. Nodes represent random variables and arcs between nodes (linking parent nodes to child nodes) indicate causal or influential relationships. Typically, discrete states are associated to each node (as an example, Fig. 2 anticipates the CESA-Q BBN). The quantitative relationships between the nodes are represented by conditional probabilities: each outcome (state) of the child node has a conditional probability given each combination of the states of the parent nodes. For example, with reference to Fig. 2, a conditional probability distribution covering all states of child node "Verification (cognitive)" is required for all possible combinations of the states of the corresponding parent nodes "Verification Hints", "Verification Means", "Verification Difficulty".

The primary hurdle for BBN development is the quantification of its CPDs. The main reason is that, for each parent node, the CPD number grows exponentially on the number of child states, thus becoming quickly large (as an example, the BBN in Fig. 2 requires the knowledge of $297=5^3+5\times 2^2+3^2+5^3$ CPDs). Data-rich applications such as medical diagnosis and financial applications typically rely on data to build CPDs by learning algorithms, also able to deal with possibly incomplete data (indeed data may not be available for all CPDs) [18]. For rare-event applications, BBNs are typically constructed based on input from domain expert, via questionnaires, interviews and panel discussions. Again, the challenge is to quantify the large number of CPDs from small data sets and expert judgment. The development of filling-up methods, or more in general, methods to determine CPDs from limited information is an important subject of research for BBNs, motivated either because of the difficulty to collect statistically significant data covering all BBN relationships, or because the elicitation of an excessive number of probabilities may become impractical and prone to biases and inconsistencies [33]. A recent review of five fill-up methods has shown that the functional interpolation shows the largest modeling flexibility [35]. However, the information requirements for this method grow exponentially with the model size (the number of BBN nodes), thus hindering the application of method for large BBN models. For the CESA-Q application presented in this paper, this did not pose difficulties,

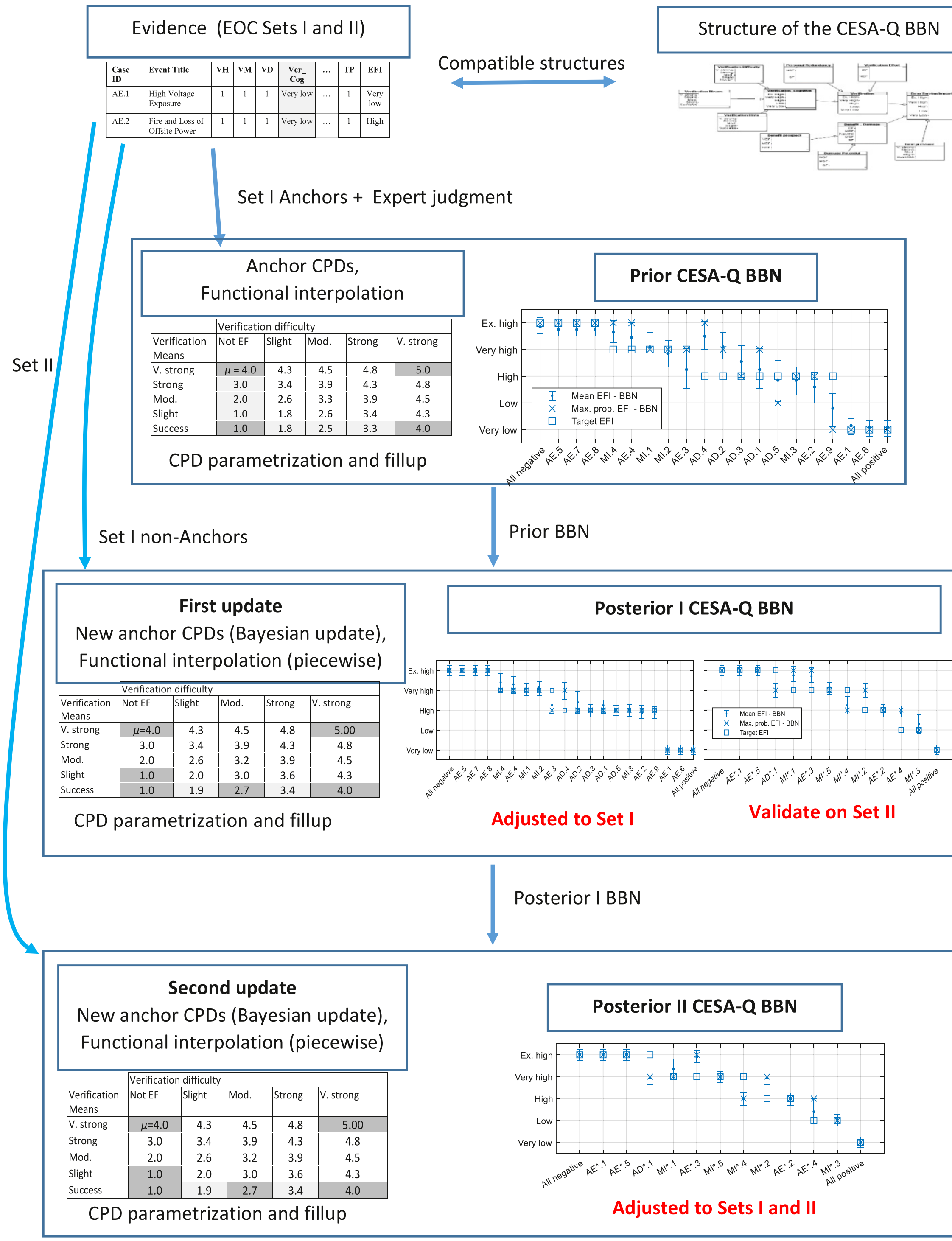

Fig. 1. Process for the development of the CESA-Q BBN (tables and figures above are explained in details in the rest of the paper).

due to the relatively small size of the BBN model (Fig. 2).

The two EOC Sets I and II (Section 2) represent the main empirical source for the BBN development (Fig. 1, top left). The sets provide patterns of adjustment factor ratings and corresponding error-forcing impact to be used to determine the BBN CPDs. The small dimension of the sets does not allow for a fully data-driven BBN development. The BBN building process devised in this paper is intended to develop the BBN around these examples, combining the functional interpolation method and expert judgment. In particular, the CPD filling algorithm populates the BBN from knowledge of the CPDs at specific, so-called "anchor" combinations of the BBN factor states: these anchoring CPDs (i.e. the CPDs corresponding to anchor factor state combinations) are derived from the EOCs sets, when available, otherwise expert judgment (Fig. 1, Box "Anchor CPDs, Functional interpolation").

This first application of the algorithm results in a Prior CESA-Q BBN (Fig. 1, Box "Prior CESA-Q BBN"). Indeed, the EOC sets include as well factor rating combinations that correspond to non-anchor CPDs: these are used to inform the corresponding CPDs, interpreting the EOC sets as new evidence ("new anchors") and updating the CPDs through a Bayesian parameter estimation (Fig. 1, Box "First update"). As shown in

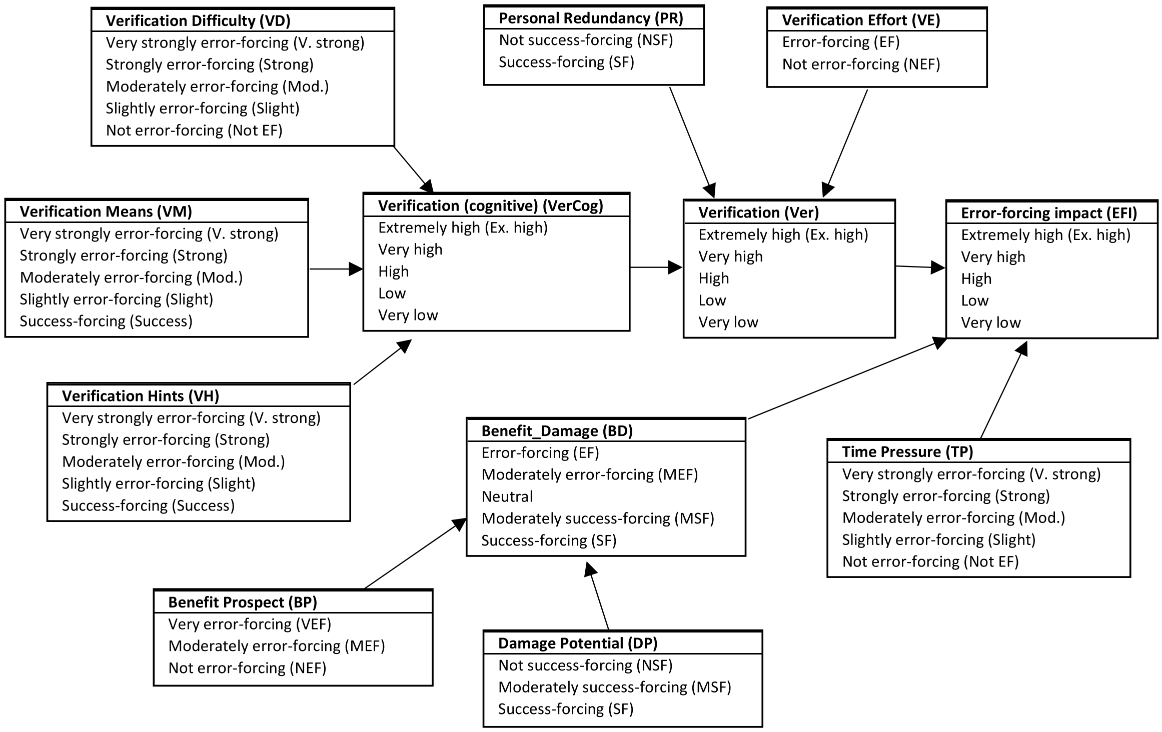

Fig. 2. Structure and node states of the CESA-Q BBN (abbreviations given in parenthesis, when needed).

Fig. 1, and discussed in more detail below in Section 3.3, the updates result in two posterior CESA-Q BBNs: Posterior I, updated based on the evidence from EOC Set I and Posterior II, obtained updating Posterior I based on evidence from EOC Set II. Indeed, mathematically, Posterior II can be obtained directly from the Prior CESA-Q BBN, aggregating the evidence from both EOC sets. However, the corresponding update was kept separate to allow using EOC Set II for validation of the Posterior I BBN. Then, Set II is again used to update the CPDs (Fig. 1, Box "Second update"), to further adjust the BBN predictions to a larger set of events as well as to generalize to analyst-to-analyst differences between Set I and Set II (see Section 2). These aspects will be discussed in more details in Section 3.3. The rest of this section provides more details on the BBN development process.

3.1. Structure of the CESA-Q BBN and of the evidence (EOC Sets I and II)

The process starts with the development of the CESA-Q BBN structure, the definition of its nodes and states and the node connections. Concerning the BBN input nodes, the most natural choice has been to maintain the CESA-Q adjustment factors and levels, see Fig. 2 for the result: the eight adjustment factors become the eight BBN input nodes, with BBN factor states defined along the adjustment factor scale (see Table A1). It is worthwhile noting that the BBN only includes CESA-Q adjustment factors and the situational features do not appear; indeed, situational features are used in CESA-Q to characterize qualitatively the situation drivers, while quantification of the decision failure probability (conditional on the situation) is done by applying the adjustment factors.

The BBN structure in Fig. 2 is developed based on the authors' judgment, with the introduction of new, intermediate nodes grouping together factors with similar effects on the error-forcing influence. For example, node "Verification (cognitive)" groups together the adjustments factors related to the cognitive elements of verification that the decision is inappropriate (e.g., presence and quality of cues, cognitive complexity). Node "Benefit_Damage" aggregates the factors addressing incentives to, or not to, make the specific decision. The fundamental benefit of the factor aggregation is the large decrease in the CPDs, one for each possible parent state combination (about 22,000 if all factors are disaggregated, $279=5^3+5 \times 2^2+3^2+5^3$ for the developed BBN in Fig. 2. The states of the intermediate nodes are given in Fig. 2, defined in line with the CESA-Q EFI scale (Fig. 2 also gives abbreviations).

Once the BBN structure is completed, each EOC set entry (Tables A2 and A3) needs to be adapted to be used as evidence for the CPDs. Indeed, the CPDs relate parent nodes and child nodes (e.g., VH, VM, VD to VerCog in Fig. 2), while the EOC sets relate the CESA-Q adjustment factors directly to the EFI. As shown in Tables A2 and A3, each EOC case has been complemented with the applicable states for each of the three intermediate nodes. The selection of the applicable states was done by the authors of the present paper after review of the original analyses in [39,40]. With the complemented set, each entry brings three pieces of evidence, one for each of the three intermediate nodes ("Verification (cognitive)", "Verification", "EFI"). For each intermediate node, the evidence applies to the combinations of the parent states matching the EOC case evaluations. For example, for node VerCog, event AE.5 carries the following pieces of evidence (see Table A2 for the event and Table A1 for the score coding):

$$\mathrm{VerCog}(\mathrm{VH}=\mathrm{V.strong},\ \mathrm{VM}=\mathrm{Mod.},\ \mathrm{VD}=\mathrm{Mod.})=\mathrm{Ex.high}$$

The evidence is then used to inform the anchor CPDs, when the factor rating combinations match the anchor combinations required by the interpolation algorithm. Otherwise, they are used in the update process to obtain Posterior BBN I and II.

3.2. Anchor CPDs and fillup by algorithm: prior CESA-Q BBN

The functional interpolation method populates the missing CPDs in three steps [35]. First, the so-called anchor CPDs are approximated by functions, described by parameters (for the present work, the mean μ and standard deviation σ of the Normal function). Second, the parameters of the full set of approximating functions are obtained by interpolating among the anchor ones. In the last step, the approximating functions are then discretized back to obtain the full set of CPDs. Example applications of the algorithm are presented in Section 4.1.

The anchor CPDs consist of all combinations of parent states at their

Table 2
Distribution for representation of uncertain evidence. Probability that evidence is assessed as represented by one state (e.g., "Very low" on one row), given different real states (each column of the "Very low" row).

		Real state of evidence				
		Very Low	Low	High	Very high	Ex. high
Assessed state for evidence	**Very low**	0.9[(1)]	0.1	1e-3	1e-3	1e-3
	Low	0.1	0.8	0.1	1e-3	1e-3
	High	1e-3	0.1	0.8	0.1	1e-3
	Very high	1e-3	1e-3	0.1	0.8	0.1
	Ex. high	1e-3	1e-3	1e-3	0.1	0.9

(1) Values in columns do not sum to 1 because of approximation (applies to all columns of the table).

lowest and highest states. For example, node VerCog requires eight ($2^3 = 8$) CPDs, one for each combination of the three parent nodes. In total, the development of the prior BBN requires the quantification of 28 anchor CPDs ($2^3 = 8$ for VerCog, $2^3 = 8$ for Ver, $2^2 = 4$ for Benefit_Damage, and $2^3 = 8$ for EFI). The identification of the EOC set entries relevant to anchor CPDs and the quantification of the expert-based anchors are presented in Section 4.1.

Given the small size of the EOC Set I, the conversion of evidence from Table A2 into anchor CPDs cannot be done via statistical analysis. Therefore, the child node state corresponding to a particular combination is interpreted as the most likely state, with some likelihood given to the other states as well. For example, for anchor combination VH=VM=VD="Success", EOC case AE.1 provides evidence of VerCog="Very low" (Table A2). This evidence is converted into about 90% probability for state "Very low", 10% probability for state "Low", and some residuals (0.1%) for the other states. The general thrust for these values is to give about 10% likelihood that the evidence is incorrect by one level (e.g., "Low" instead of "Very Low") in both directions.

The parameters of the approximating functions are obtained by minimizing the sum of the squared difference between each value of the anchor CPDs and the corresponding value of the Normal function (appropriately normalized so that the values sum to 1). The Normal function is defined on an underlying continuous scale: the scale ranges from $-\infty$ to $+\infty$ with the parent node states corresponding to integer values above 1. For example, for parent "Verification Difficulty", the "Not EF", "Slight", "Mod.", "Strong" and "V. Strong" states correspond to the values of 1, 2, 3, 4 and 5, respectively (similarly, for the other underlying factor scales). This underlying scale is only used for the interpolation, and, given the linearity of the interpolation across the states, has no effect on the final CPDs produced. The reader can refer to [35] for more details.

3.3. Bayesian update of the CPDs and derivation of posterior I and II CESA-Q BBNs

Once the Prior BBN is built, the actual entries of EOC Set I (or II in the subsequent stage) are used to adapt the BBN response to the additional evidence. The process for doing this is the same for both EOC Sets I and II. The idea is to apply the classical Bayesian estimation update to selected CPDs of the Prior BBN, based on the evidence from the EOC set.

For each EOC set entry, and for each of the intermediate nodes, the CPDs of the parent state combinations corresponding to the entry are selected. The CPDs of the Prior BBN become themselves prior information about the child state, to be updated based on the evidence. As an example, take EOC event AE.9 from Table A2, for node "Verification (cognitive)". The EOC entry brings the evidence of "Verification (cognitive)"="High", for the parent state combination VH="Success-forcing" (score 1 in Table A2), VM="Success-forcing" (score 1), VD="Moderately error-forcing" (score 0.5).

The update itself is done according to the Bayesian estimation model for uncertain evidence presented originally by [41] and specified for human reliability applications by [42]. The observed evidence (e.g., "Verification (cognitive)"="High") is interpreted as a random variable, associated itself an uncertainty distribution which models variability in the evidence and in its interpretation. The assumption is that, for a particular observation, a true value exists (i.e. an applicable parent state), but the assessed evidence may be different, e.g., due to subjectivity of the interpretation. For the work presented in this paper, the uncertainty distribution represents the probability that the evidence is interpreted as one state (e.g., "High"), given the real state is any of the other ones. The distribution is reported in Table 2: for each of the assessed states, the distribution is centered on the real ones, therefore representing an unbiased estimate. Then, about 10% probability is associated to incorrect estimation by one state (e.g., assessed as "High", when the real state would be "Very High"). Larger deviations are assigned probability of 0.001. The specific values were chosen to allow comparable weight with the prior distributions at the anchor positions and resulting prior CPDs (see Section 3.2 and uncertainty distributions in Fig. 5). In the Bayesian update process, the likelihood of the evidence is then calculated by combining the probability that the child node would manifest as one of its possible states (i.e. the CPD), with the probability of observing the particular evidence, as presented in [41,42]. See Section 4.2 for an example update.

It is worthwhile mentioning the difference between this implementation of the Bayesian update and a similar one from [43], also recently developed to incorporate new empirical evidence into prior knowledge about human error probabilities. In particular, the most interesting difference is in the type and, consequently, in the appearance of the evidence. The present case addresses operational events, which constitute single manifestations of failures in specific contexts, as explained in Ch. 3.1. The evidence from each event (the corresponding EFI) is then associated a distribution to model variability and uncertainty of the evidence itself, as presented above and in more details in [42]. In the study from [43], evidence is collected from simulator data, representing multiple manifestations of the same context, including failures and successes by different operating crews. Therefore, evidence comes in the form of number of failures over number of successes and the evidence uncertainty is related to the statistical uncertainly due to the sample size.

The evidence from the EOC set is not only used to update the corresponding CPDs, but can be used to adapt the BBN response at large, also adapting the CPDs of neighboring state combinations. This is important especially in case the difference in the EOC set evidence and the Prior BBN response is very large: this may suggest that the linear interpolation along all parent states does not represent the actual behavior. This step is also important because the change of single CPDs may introduce non-coherent BBN behavior (e.g., if one input becomes more error-forcing, the BBN output EFI state cannot become less error-forcing). Of course, large differences in the EOC set evidence and the Prior BBN response may also suggest the need to revisit the expert judgment based anchors, such that the Prior BBN reflects better the whole evidence in the first place. Indeed some iteration in the definition of the anchors and the BBN response over the whole EOC set may often be necessary (as it was for the present work). Operatively, the adaptation of the BBN to the updated CPDs is then done by converting the posterior CPDs to Normal approximating functions (the parameters μ an σ of each function are identifying by least square errors between the approximation and the CPD). Then, the new values of the μ's are used as the additional anchor values, and the interpolation operates piecewise including the original and additional anchors. For the σ's, only the ones relevant to the EOC set are updated, while all others are kept at their prior value. The next Section 4.2 provides an example concerning "Verification (cognitive)".

Table 3
Anchors for intermediate node "Verification (cognitive)": combinations of the parent node states, corresponding evidence and basis.

	Anchor combinations of the parent node states			Evidence for CPD of VerCog	
Anchor #	VH[1]	VM	VD	State	Basis[2]
1	V. strong	V. strong	V. strong	Ex. high	AE.8
2	V. strong	V. strong	Not EF	Ex. high	Expert Judgment
3	V. strong	Success	V. strong	Ex. high	Expert Judgment
4	V. strong	Success	Not EF	Ex. high	Expert Judgment
5	Success	V. strong	V. strong	Ex. high	Expert Judgment
6	Success	V. strong	Not EF	Very high	Expert Judgment
7	Success	Success	V. strong	Very high	Expert Judgment
8	Success	Success	Not EF	Very low	AE.2, AE.1, AE.6
8*[3]	Slight	Success	Not EF	Very low	Expert Judgment
9*[3]	Success	Slight	Not EF	Very low	Expert Judgment

[1] Abbreviations for nodes and states given in Fig. 2.
[2] Basis: ID from EOC Set I (see Table A2) or Expert Judgment.
[3] Anchors added to incorporate success-forcing effect of "Verification Means" and "Verification Hints" in their most positive state (see definition in Table A1).

4. BBN development implementation

4.1. Prior CESA-Q BBN

For simplicity of the presentation, the details of the BBN development are given with reference to the intermediate node "Verification (cognitive)", which aggregates the cognitive aspects connected with the verification of the decision. All parameters μ and σ and the corresponding CPDs of the developed BBNs are available as supplementary material on the paper web page.

The anchors required for the "Verification (cognitive)" intermediate node are reported in Table 3. The table shows that two of the anchors are based on EOC Set I, both relating to all factor states at the same extreme values (either all with negative or positive influence, Anchors 1 and 8, respectively). The remaining, judgment-based, anchors were derived based on the following considerations:

- If "Verification Hints" is in its most negative state ("V. strong"), then "Verification (cognitive)" is in its "Extremely high" (error-forcing impact) state, independently on the state of the other factors (Anchors 2–4).
- When "Verification Hints" is in its most positive state ("Success"):

  ○ if both other factors have their strongest negative influence, then "Verification (cognitive)" is in its "Extremely high" (error-forcing impact) state (Anchor 5).
  ○ If only one of the other two factors is in its most negative state, then "Verification (cognitive)" is in its "Very high" (error-forcing impact) state (Anchors 6, 7).

The above rules reflect the strong importance of "Verification Hints" to result in an error-forcing condition. Given the lack of hints to verify

Table 4
Prior CESA-Q BBN, node "Verification (cognitive)": values of the μ parameter of the CPD approximating functions.

		Verification difficulty (VD)				
Verification Hints (VH)	Verification Means (VM)	Not EF	Slight	Mod.	Strong	V. strong
V. strong	V. strong	5.00 (2, EJ)$^{(1, 2)}$	5.00	5.00	5.00	5.00 (1, AE.8)
	Strong	5.00	5.00	5.00	5.00	5.00
	Mod.	5.00	5.00	5.00 (AE.5, 5)$^{(3)}$	5.00	5.00
	Slight	5.00	5.00	5.00	5.00	5.00
	Success	5.00 (4, EJ)	5.00	5.00	5.00	5.00 (3, EJ)
Strong	V. strong	4.75	4.81	4.88	4.94	5.00 (AE.7, 5)
	Strong	4.48	4.59	4.71 (MI.4, 4)	4.82	4.94
	Mod.	4.21	4.38	4.54	4.71	4.88
	Slight	3.94	4.16	4.38	4.59	4.81
	Success	3.67 Data$^{(4)}$	3.94	4.21 (MI.1, 4)	4.48	4.75
Mod.	V. strong	4.50	4.63	4.75	4.88	5.00
	Strong	3.96	4.19	4.42	4.65	4.88
	Mod.	3.42	3.75	4.08 (AD.2, 3)	4.42	4.75
	Slight	2.88	3.31	3.75	4.19	4.63
	Success	2.33	2.88	3.42 Data$^{(5)}$	3.96	4.50
Slight	V. strong	4.25	4.44	4.63	4.81	5.00
	Strong	3.44	3.78	4.13	4.47	4.81
	Mod.	2.63	3.13	3.63	4.13	4.63
	Slight	1.81	2.47	3.13	3.78	4.44
	Success	1.00 (8*, EJ)	1.81	2.63 (MI.3, 3)	3.44	4.25
Success	V. strong	4.00 (6, EJ)	4.25	4.50	4.75	5.00 (5, EJ)
	Strong	3.00$^{(6)}$	3.44	3.88	4.31	4.75
	Mod.	2.00	2.63	3.25 (AE.4, 3)	3.88	4.50
	Slight	1.00 (9*, EJ)	1.81	2.63	3.44	4.25
	Success	1.00 (8, Data$^{(7)}$)	1.75	2.50 (AE.9, 3)	3.25	4.00 (7, EJ)

(1) Scale for parameter μ: 1 - Very low, 2 – Low, 3 – High, 4 – Very high, 5 – Ex. high.
(2) Dark gray cells identify anchor combinations of parent states. In parenthesis: (anchor # from Table 3, basis for anchor: expert judgment, EJ, or EOC Set I event).
(3) (ID event, value) outside dark gray cells identifies non-anchor data from EOC Set I and corresponding VerCog value on scale$^{(1)}$.
(4) Multiple entries: (AD.1, 3), (AD.4, 3), (AE.3, 4).
(5) Multiple entries: (AD.3, 3), (AD.5, 3), (MI.2, 3).
(6) Light gray cells show examples of interpolation directions. Along VM: 1.0, 2.0, 3.0, 4.0; along VD: 1.00, 1.75, 2.50, 3.25, 4.00.
(7) Multiple entries: (AE.1, 1), (AE.2. 1), (AE.6, 1).

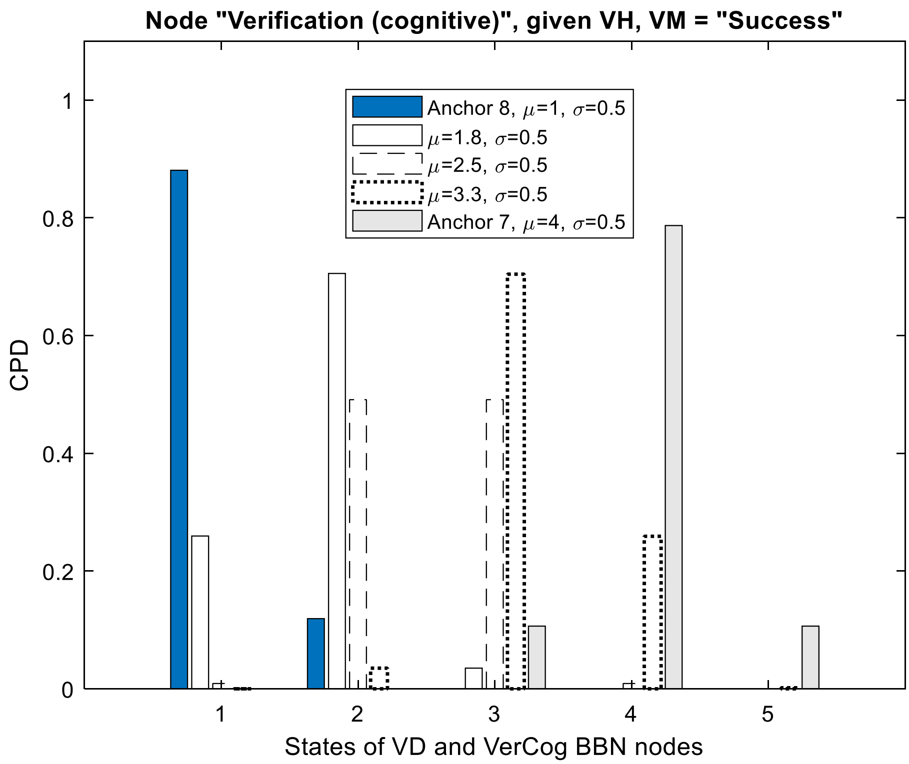

Fig. 3. Quantification of the CPDs via functional interpolation algorithm. Example for "Verification (cognitive), given VH = "Success", VM = "Success". The figure shows the five CPDs corresponding to the different states of VD. Filled bars: anchors; Empty bars: interpolated. Numerical scale for VerCog: 1 - Very low, 2 – Low, 3 – High, 4 – Very high, 5 – Ex. High; numerical scale for VD: 1 – Not EF, 2 – Slight, 3 – Mod., 4 – Strong, 5 – V. Strong.

the decision (see VH definition in Table A1), the EOC is deemed as certain.

Note from Table 3 the two additional anchors 8* and 9*, positioned in non-extreme node states (nodes VH and VM at "Slightly error-forcing" for anchors 8* and 9*, respectively). These anchors were added to reflect the success-forcing effect of VH and VM at their most positive state ("Success-forcing", Table A1). Indeed, as shown in Table 3, both Anchors 8* and 9* force the "Verification (cognitive)" state to "Very low", compensating the slightly error-forcing effect of one factor with the success-forcing effect of the other.

The anchor evidence from Table 3 is converted into CPD approximating functions (Normal functions), with the parameters μ corresponding to the numerical value on the 1 to 5 scale (for the five states of "Verification (cognitive)" and σ equal to 0.5. The μ values for the anchor combinations are reported in Table 4 in the dark gray cells. The value of 0.5 for σ was determined to allow reasonable uncertainty on the anchor judgment. Fig. 3 shows the effect for node "Verification (cognitive)", given VH = "Success" and VM = "Success". The figure shows the two Anchors 7 and 8 (filled bars in the figure, at VD = "Very high", value 4 in Figure, and VD = "Very low", value 1 in Figure, respectively), with corresponding parameter μ equal to 4 and 1, respectively (Tables 3 and 4). As shown by the two anchor CPDs, the uncertainty set by σ=0.5 corresponds to about 0.1 probability to the level below or above the corresponding one.

The results of the application of the interpolation algorithm are presented in Table 4, concerning the values of the μ parameter of the CPD approximating functions for node "Verification (cognitive)". The table shows the interpolation along all the input factor directions: e.g., light shaded cells show interpolations along Verification Difficulty, with the values 1.00 (at anchor 8), 1.75, 2.50, 3.25, 4.00 (at anchor 7) and along Verification Means, with the values 1.00 (at anchor 9*), 2.00, 3.00, 4.00 (at anchor 6). Fig. 3 shows the CPD results along the VD direction, exemplifying the gradual shift of the CPDs within the constraining anchors.

The definition of the anchors and the application of the algorithm was done in similar way for the other nodes "Verification" and "EFI". For node "Verification", the interpolation operates only along the direction of "Verification (cognitive)", because the other two nodes are binary ("Personal Redundancy" and "Verification Effort"), therefore only anchor CPDs exist. For node "EFI", the interpolation along the direction of node "Verification" is performed with an additional constraint resulting from the definitions of the states of the two other nodes entering "EFI". In particular, the combination of factors and states "Benefit_Damage"="Neutral" and "Time Pressure"="Not Error-Forcing" should have no effect on the error-forcing impact in addition to the effect of node "Verification" alone. This is why for that combination of states the values of parameter μ for EFI match the values for node "Verification". Finally, for node "Benefit_Damage", the whole set of CPDs was quantified with expert judgment because of the small number of CPDs. This node was not subjected to the subsequent updates (all CPDs and parameters of the approximating functions are included as the supplementary material on the paper web page).

With the whole set of CPDs completed, the Prior CESA-Q BBN is then applied to the EOC cases of Set I, to adjust its output to the available evidence also for non-anchor factor combinations. Fig. 4 (top) shows the response of the Prior CESA-Q BBN on the Set I analysis cases ("□" indicate the target EFI, "x" indicate EFI state with highest likelihood according to the BBN). The events in the figure are sorted by the EFI (obtained by converting the reliability index assigned to each of them in [38]), and then by EFI prediction by the BBN (highest to lowest). Therefore, the generally decreasing trend of the BBN predictions suggests its overall ability to discriminate across different difficulty levels. In four cases, the CESA-Q BBN bounds (50% confidence bounds) do not cover the target EFI assignment, in line with the expectation from a 50% confidence interval. In seven cases, the most likely EFI level for the BBN does not match the target EFI. It has to be noted that none of these cases was used as anchor evidence; therefore some mismatch in the predictions is not surprising.

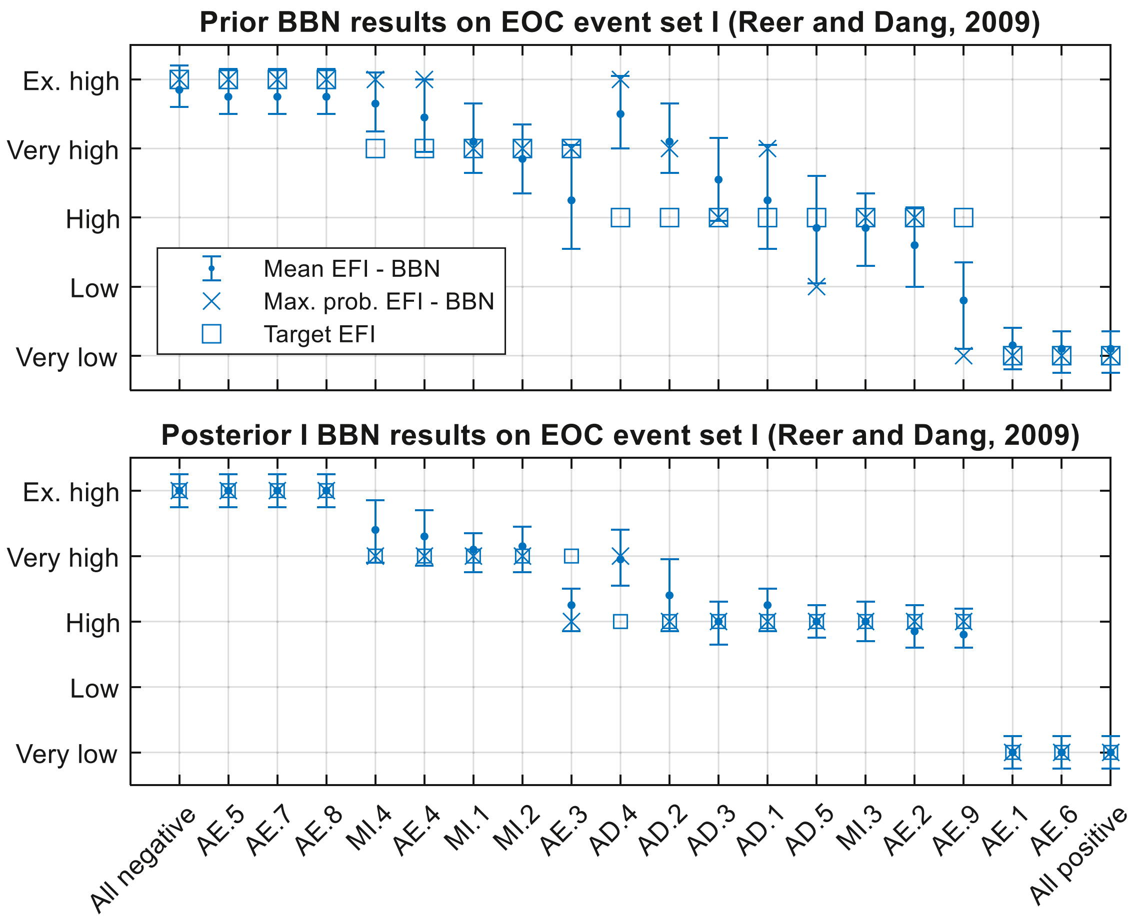

Fig. 4. Comparison of the CESA-Q BBN results vs. error forcing impact from EOC Set I [39]. Events ordered by target EFI (Table A2), then by mean EFI from CESA-Q BBN. Top: Prior BBN, bottom: Posterior I BBN. Error bars identify 50% confidence bounds.

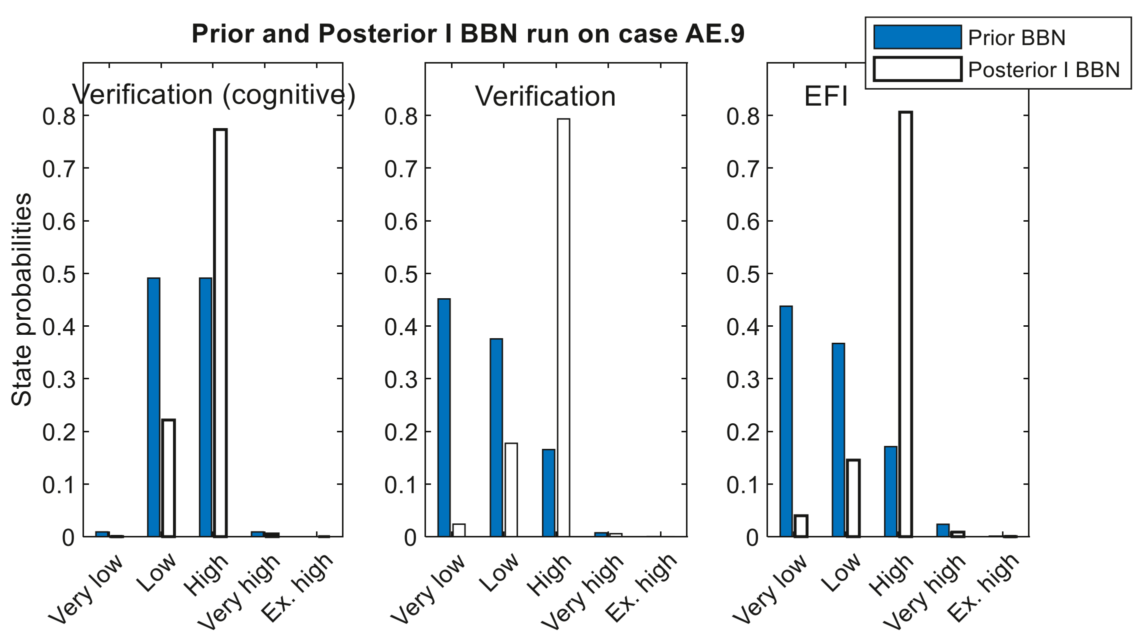

Fig. 5. Assessment of the CESA-Q BBN on one EOC Set I case: AE.9, on nodes "Verification (cognitive)", "Verification" and "Error-forcing impact". Filled bars: Prior BBN, empty bars: Posterior I BBN. AgenaRisk [44] software used for all BBN runs.

In Fig. 4 (top) a relatively strong mismatch can be seen for event AD.4, at the break in the trend that can be seen in the center of the figure. The reason for the mismatch can be understood by comparing AE.3 and AD.4. Based on the target EFI, AE.3 has a higher EFI (higher error forcing index or probability of occurrence) than AD.4. The prior BBN model instead predicts a higher EFI for AD.4. Examining the

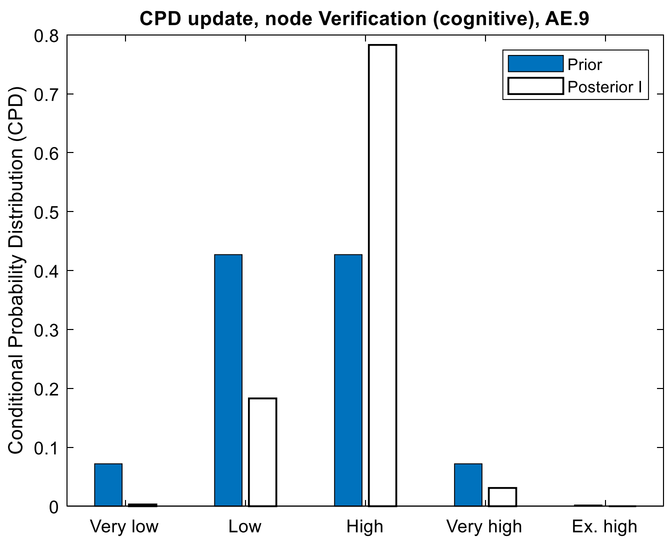

Fig. 6. Example update of CPDs for event AE.9, node "Verification (Cognitive)", given VH = "Success", VM = "Success", VD="Moderate". Approximating function parameters: Prior $\mu = 2.5$, $\sigma = 0.5$; Posterior I $\mu = 2.7$, $\sigma = 0.4$.

adjustment factor evaluations for the two events that underlie the target EFIs in Table A2, the adjustments factors are slightly better for AE.3 than for AD.4 (values between 0 and 1, larger values are reliability-enhancing or more favorable) while the more holistic target EFI assignments indicate that AE.3 is more error-forcing. This mismatch is not the effect of an inconsistency within EOC Set I. It is worth noting that the two events AE.3 and AD.4 belong to different types of error-forcing situations (Adverse Exception and Adverse Distraction, respectively); consequently, a single function (the BBN) may not adequately represent the effect of the adjustment factors for the two types of situations. A potential solution would be to develop multiple models rather than a single model; retaining a single BBN for all cases was preferred for two reasons. First, this issue arose only for one case. Second, the multiple models would each have considerably weaker data supports while requiring additional anchor evaluations.

The origin of the mismatches between EOC Set I and the Prior CESA-Q BBN can be traced by taking, as an example, case AE.9. For AE.9, the target EFI value is "High", while the Prior BBN returns lower values (mean value around "Low", most likely value "Very low", see also Fig. 4, top). The Prior BBN response on nodes "Verification (cognitive)", "Verification" and "Error-forcing impact" is shown in Fig. 5 (filled bars). Taking node "Verification (cognitive)" as example, the BBN response gives equal probabilities to the two states "High" and "Low". This assessment is the result of the CPD for the corresponding factor combination (VH= Success, VM= Success, VD= Mod.). The CPD, in turn, is the result of the parameter $\mu = 2.5$ (Table 4). Still from Table 4, it can be seen that the EOC Set I value for "Verification (cognitive)" AE.9 is 3 ("High"). The underestimation proceeds along the subsequent node of the BBN ("Verification"), producing an overall underestimated EFI value (last graph on the right in Fig. 5, filled bars). In Section 4.2, EOC Set I is used to update the CPDs of the BBN, adding this evidence to the prior knowledge.

Table 5
. Posterior I CESA-Q BBN, node "Verification (cognitive)": values of the μ parameter of the CPD approximating functions.

		Verification difficulty				
Verification Hints	Verification Means	Not EF	Slight	Mod.	Strong	V. strong
V. strong	V. strong	5.00[1, 2]	5.00	5.00	5.00	5.00 (AE.8, 5)
	Strong	5.00	5.00	5.00	5.00	5.00
	Mod.	5.00	5.00	5.00 (AE.5, 5)[3]	5.00	5.00
	Slight	5.00	5.00	5.00	5.00	5.00
	Success	5.00	5.00	5.00	5.00	5.00
Strong	V. strong	4.75	4.81	4.88	4.94	5.00 (AE.7, 5)
	Strong	4.44	4.56	4.69	4.81	4.94
	Mod.	4.13	4.26	4.40 (MI.4, 4)	4.64	4.88
	Slight	3.81	4.08	4.35	4.58	4.81
	Success	3.50 Data[4]	3.90	4.30 (MI.1,4)	4.53	4.75
Mod.	V. strong	4.50	4.63	4.75	4.88	5.00
	Strong	3.96	4.19	4.42	4.65	4.88
	Mod.	3.42	3.51	3.60 (AD.2, 3)	4.18	4.75
	Slight	2.88	3.11	3.35	3.99	4.63
	Success	2.33	2.72	3.10 Data[5]	3.80	4.50
Slight	V. strong	4.25	4.44	4.63	4.81	5.00
	Strong	3.44	3.78	4.13	4.47	4.81
	Mod.	2.63	3.13	3.63	4.13	4.63
	Slight	1.81	2.47	3.13	3.78	4.44
	Success	1.00	1.90	2.80 (MI.3, 3)	3.53	4.25
Success	V. strong	4.00	4.25	4.50	4.75	5.00
	Strong	3.00	3.44	3.88	4.31	4.75
	Mod.	2.00	2.60	3.20 (AE.4, 3)	3.85	4.50
	Slight	1.00	1.98	2.95	3.60	4.25
	Success	1.00 Data[6]	1.85	2.70 (AE.9, 3)	3.35	4.00

(1) Scale for parameter μ: 1 - Very low, 2 – Low, 3 – High, 4 – Very high, 5 – Ex. high.
(2)Dark gray cells identify anchor combinations of parent states.
(3)(ID event, value) outside dark cells identifies non-anchor data from EOC Set I and corresponding VerCog value on scale[(1)].
(4)Multiple entries: (AD.1, 3), (AD.4, 3), (AE.3, 4).
(5)Multiple entries: (AD.3, 3), (AD.5, 3), (MI.2, 3).
(6)Multiple entries: (AE.1, 1), (AE.2. 1), (AE.6, 1).

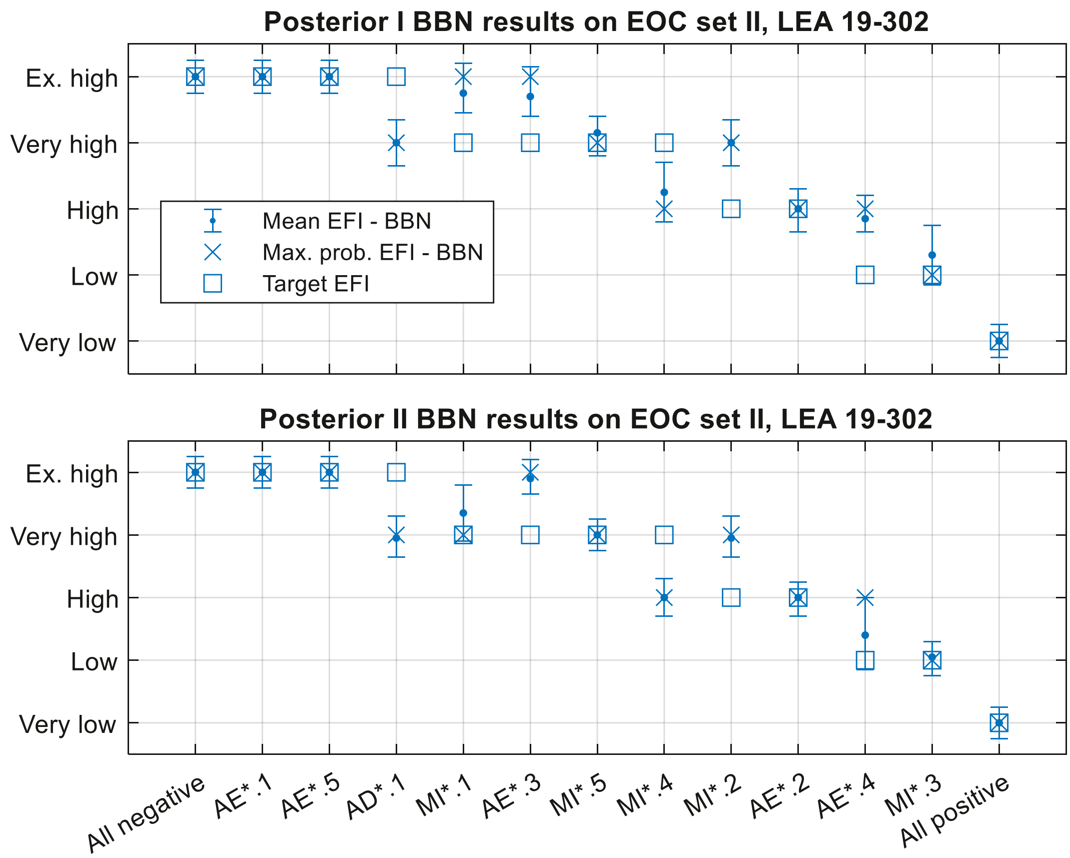

Fig. 7. Comparison of the CESA-Q BBN results vs. error forcing impact from EOC Set I [40]. Events ordered by target EFI (Table A3), then by mean EFI from CESA-Q BBN. Top: Posterior I BBN, bottom: Posterior II BBN. Error bars identify 50% confidence bounds.

4.2. Posterior I CESA-Q BBN

The process for the update to Posterior I is presented still with reference to EOC event AE.9. For node “Verification (cognitive)”, the EOC entry brings the evidence of “Verification (cognitive)”=”High”, for the parent state combination VH=Success, VM=Success, VD=Mod. (Table A2). Fig. 6 shows the effect of the evidence on the CPDs, which after the update of the parameters μ and σ correspondingly shifts mass towards the evidence (see also the increase in parameter μ from 2.5 in the Prior to the value 2.7 of Posterior I). Fig. 5 (empty bars) shows the effect of the updated CPDs at “Verification (cognitive)” and at “Verification” to result in EFI distributed closer to the target value “High”.

The updated values of parameter μ of the approximating functions become new anchors for interpolation. In Table 5, the μ values in correspondence of EOC events, (ID event, value) in Table 5 are the result of the update. The rest of the values are obtained via piecewise interpolation across the original and the newly determined anchors. The last row of Table 5 shows an example: the new anchor μ = 2.7 becomes an additional interpolation point and the other two intermediate values (corresponding to VD= Slight and VD = Strong) are obtained as the averages of the two neighboring cells. Table 5 also shows instances of the need to repeat with the interpolation after the update, because of possible non-coherent behavior of the posterior BBN. See the cell in Table 5 right above the one corresponding to AE.9 (VH=Success, VM=Slight, VD=Moderate), with μ = 2.95 (Posterior I) and μ = 2.63 (Prior, Table 4). If the prior value would not be recalculated, moving from this cell down to the AE.9 cell would result in an increase of error forcing impact (to 2.7, Table 5) while factor VH becomes less error forcing (from “Slight” to “Success”), thus representing a non-coherent behavior.

Fig. 4 (bottom) shows the Posterior I BBN response on EOC Set I. The effect of the update to reproduce better the target EFI impact is evident. As for the Prior BBN, the decreasing trend interrupts at AD.4, due to the imposed coherent response of the BBN. The smaller interruptions of the decreasing trend (e.g., between MI.1 and MI.2) appear because, to ease comparison of the results, the events in Fig. 4 (bottom) are kept in the same order as in Fig. 4 (top): as an effect of the update, the EFI mean also changes. At this stage of the process in Fig. 1, Posterior I BBN represents a model that acceptably reproduces EOC Set I. As a next step in the process, the BBN is tested on EOC Set II, which has not yet been used for the BBN development.

4.3. Posterior II CESA-Q BBN

The response of Posterior I BBN on EOC Set II is shown in Fig. 7 (top). For five out of the eleven EOC events in the set, the BBN prediction interval (50% confidence) covers the target EFI values. In these cases, the most likely EFI value as per the BBN prediction matches the target value. In the remaining six cases, the BBN predictions are off by one EFI level at most. The fact that no large differences exist (e.g., EFI predictions off by two or more levels) is also visible by the generally decreasing trend of the BBN predictions: similar to the considerations in Sections 4.1, 4.2 this suggests the general ability of the model to discriminate the different levels of error-forcing impact over their possible spectrum. Fig. 7 (top) shows acceptable validity of the Posterior I BBN to reproduce the assessments in EOC Set II.

The final stage for the CESA-Q BBN development is to update the model response to EOC Set II, in the same way as done for the first update to Set I. The final set of CPD approximating function parameters μ is reported in Table 6. The response of the CESA-Q BBN to the EOC set is given by Fig. 7 (bottom). Overall, as expected, the response of the BBN adjusts to the evidence in the set. In terms of actual BBN predictions,

Table 6
Posterior II CESA-Q BBN, node "Verification (cognitive)": values of the μ parameter of the CPD approximating functions.

		Verification difficulty				
Verification Hints	Verification Means	Not EF	Slight	Mod.	Strong	V. strong
V. strong	V. strong	5.00(1, 2, 3)	5.00	5.00	5.00	5.00
	Strong	5.00	5.00	5.00	5.00	5.00
	Mod.	5.00	5.00	5.00 (AE.5, 5)(4)	5.00	5.00
	Slight	5.00	5.00	5.00	5.00	5.00
	Success	5.00	5.00	5.00	5.00	5.00
Strong	V. strong	4.40 (MI*.1, 4)	4.55	4.70	4.85	5.00 (AE.7, 5)
	Strong	4.18	4.36	4.55	4.76	4.98
	Mod.	3.95 (MI*.5, 4)	4.18	4.40 (MI.4, 4)	4.68	4.95
	Slight	3.73	4.01	4.30	4.61	4.93
	Success	3.50 Data(5)	3.85	4.20 Data(6)	4.55	4.90 (AE*.3, 5)
Mod.	V. strong	4.27	4.45	4.63	4.82	5.00
	Strong	3.73	3.92	4.12	4.51	4.90
	Mod.	3.18	3.39	3.60 (AD.2, 3)	4.20	4.80
	Slight	2.64	2.97	3.30	4.00	4.70
	Success	2.10 (MI*2, 2)	2.55	3.00 Data(7)	3.80	4.60
Slight	V. strong	4.13	4.35	4.57	4.78	5.00
	Strong	3.35	3.64	3.93	4.38	4.83
	Mod.	2.57	2.93	3.30 (MI*4, 3)	3.98	4.65
	Slight	1.78	2.34	2.90	3.69	4.48
	Success	1.00 (MI*.3, 5)	1.75	2.50 Data(8)	3.40	4.30
Success	V. strong	4.00	4.25	4.50	4.75	5.00
	Strong	3.00	3.44	3.88	4.31	4.75
	Mod.	2.00	2.60	3.20 (AE.4, 3)	3.85	4.50
	Slight	1.00	1.98	2.95	3.60	4.25
	Success	1.00 Data(9)	1.85	2.70 (AE.9, 3)	3.35	4.00

(1) Scale for parameter μ: 1 - Very low, 2 – Low, 3 – High, 4 – Very high, 5 – Ex. high.
(2) Dark gray cells identify anchor combinations of parent states.
(3) (ID event, value) outside dark cells identifies non-anchor data from EOC Sets I and II and corresponding VerCog value on scale(1).
(4) Light gray cells identify changes in the μ values above 0.3 (absolute) between the Posterior II and the Prior BBN.
(5) Multiple entries: (AD.1, 3), (AD.4, 3), (AE.3, 4).
(6) Multiple entries: (MI.1,4), (AD*.1, 4).
(7) Multiple entries: (AD.3, 3), (AD.5, 3), (MI.2, 3), (AE*.2, 3).
(8) Multiple entries: (MI.3, 3), AE*.4, 2).
(9) Multiple entries: (AE.1, 1), (AE.2. 1), (AE.6, 1).

Fig. 7 distinguishes three behaviors. First, new evidence in Set II reinforces the earlier belief on the EFI impact, thus decreasing the response uncertainty (see for example case MI*.3, where the BBN response narrows to the EFI of "Low"). Second, new evidence moves the earlier belief to different EFI levels, thus increasing the response uncertainty to envelop the new evidence (see for example the case of AE*.4). The last case is when multiple events provide different evidence for the same factor configuration. See for example the case of MI*.2 and MI*.4. For the same factor configuration Ver=High, BD=Neutral, VE=NEF, the evidence for node EFI for MI*.2 and MI*.4 is EFI=High and EFI=Very high, respectively (Table A3). In this last case, the resulting updated CPD considers the different evidences so that in Fig. 7 the effect on the single event is not necessarily as evident as in the first two cases (similar situations occur for AD*.1 and AE*.3).

The model underlying Posterior II represents our updated state of knowledge about CESA-Q factor influences on the EFI, based on the two EOC sets. To give some insights in the effect of the update, light gray cells in Table 6 identify changes in the μ values above 0.3 (absolute) between the Posterior II and the Prior BBN models (with reference to node "Verification (cognitive)"). The largest difference was found for the combination with all factors at their "Moderately error-forcing" state (0.48) and, in general, the area around this combination exhibits the largest differences. This is the effect of the evidences by AD.2, AD.3, AD.5, MI.2 from EOC Set I and event AE*.2 from EOC Set II. All these events suggest lower error-forcing impact for "Verification (cognitive)", compared to the one implied by the Prior BBN (see all events associated to state "High", while prior values would suggest "Very high").

5. Discussion

In this section, three important elements of the present paper are further discussed: the choice of the BBN modeling framework, the BBN development process, and the envisioned next steps to enhance the CESA-Q method.

The CESA-Q BBN constitutes an example of the modeling capabilities of BBNs and of their attractiveness for HRA applications. Adjustment factors do not have the same effect independently of the other factors. For example, saturation of the error-forcing strength is a typical effect underlying the CESA-Q model. Consider the effect of "Verification Difficulty", depending on "Verification Hints" ("Verification Means" is kept at "Success" for simplicity). Referring to Table 6, when "Verification Hints" is in its "Success" state and "Verification Difficulty" changes from its "Success" state to "V. strong", the value of μ for the child state changes of three levels (from 1 to 4, i.e., from around "Very low" to around "Very high"). When "Verification hints" is "Mod.", the same change in "Verification Difficulty" results in 2.5 levels (from 2.1 to 4.6, from around "Low" to somewhere in between "Very high" and "Ex. high". When "Verification Hints" is "Extremely high error-forcing", "Verification Difficulty" has no effect, since the first factor is enough to result in (almost) certain "Extremely high error-forcing" state for the child node.

Another recurrent effect in the CESA-Q BBN is compensation between error-forcing and success-forcing impact. As already discussed in 4.1, the two anchors 8* and 9* represent the compensatory effect of one of the two factors “Verification Hints” and “Verification Means” when in their “Success-forcing” state, when the other is “Slightly error-forcing”. Effects of compensation by “Verification Means” in its “Success-forcing” state are also visible in Anchors 6 and 7 (Table 3) when one of VM and VD is in their most negative state (VerCog state set to “Very high”). However, no compensation is modeled if both VM and VD are in their most negative state (VerCog state set to “Ex. high”, Anchor 5 Table 3).

The above effect demonstrates the flexibility of the BBN framework with respect to the typical, multiplier-based HRA models, according to which the effect (i.e. the multiplier) of one factor is independent on the state of the others. Another typical HRA modeling framework are decision trees [45,46]. Indeed, decision trees can model these dependence effects because the result of each branching combination is not necessarily bound to a simple calculation model (like the multiplicative one). Yet, decision trees are generally implemented on binary factors, thus not allowing representing graded influences as BBNs do.

Concerning the overall process for the BBN development, traceability in the use of evidence and expert judgment has been one of the driving features. Given the scarcity of the evidence, the choice has been to combine the BBN model with a Bayesian update of selected CPDs. How the EOC set evidence enters as the anchors and the subsequent updates is clearly identifiable through the process, e.g., via the various Tables 4–6. Expert judgment enters mainly in the anchor assessments. It is worthwhile to mention that the judgment is not in the form of CPDs, but in the form of the most likely corresponding level (score) of a factor. This is to avoid direct assessment of probabilities, which besides its challenges, could additionally hinder transparency (it is easier to justify and interpret a statement on the most likely state, then a statement on the probability distribution). Uncertainty on the judgement is indeed considered in the process, but externally to the judgment (i.e. via the distribution in Table 2). This eventually allows for sensitivity analysis of the model on the expert assessments. A consequence of traceability is the ease to add new evidence or additional judgment to the model. Depending on the cases, these would add to the existing anchors, or determine a new Posterior model, following the same steps presented in the papers. Indeed, the limited size of the CESA-Q BBN helps in the traceability of the development steps: for larger BBNs it would be more difficult to follow each process step and visualize it in tables such as Tables 4 and 5. On the other hand, as already mentioned, the functional interpolation method is suitable for medium-sized BBNs (e.g., 20 nodes); for much larger BBNs, a different development concept may be required, because of the combinatorial increase in information requirements by the functional interpolation algorithm.

Another worthwhile comment concerns the fact that the development process as shown in Fig. 1 foresees the use of two data sets. Having two sets is not a mathematical prerequisite, because the Bayesian update process does not depend on how the evidence is aggregated: one could obtain exactly the same result (Posterior II BBN) by updating the Prior BBN with only one set aggregating Set I and II. This allows the application of the process also in case of only one set being available as evidence, i.e. resulting directly in the final BBN, without the intermediate validation step. However, in case only one set would be available, depending on the size and structure of the set, it may still be worth splitting the set in two subsets, as common in the application of many learning algorithms, and still attempt intermediate validation.

The CESA-Q BBN links the adjustment factor ratings to the distribution of the applicable EFI. Through the EFI level-probability relationship of Table 1, this translates in a probability distribution that can be used for probabilistic safety assessment applications. Different approaches are possible for the actual conversion into a probability distribution, e.g., weighting each EFI contribution (each associated an uncertainty distribution of the error probability) or applying directly the most likely level. These aspects have not been explored by the present paper, which has focused on the BBN development. Work is ongoing to integrate the present work into the CESA-Q method [38] for EOC analysis and quantification. Still related to the conversion of the EFI levels on the probability scale, the applicability of recent data collection initiatives [47–49] is under evaluation, to validate or eventually update the current relationship (Table 1).

6. Conclusions

This paper contributes to quantification of decision-related EOCs, with a BBN model providing the error-forcing impact of a specific situation, for given ratings of the influencing factors. The model is tailored to the CESA-Q method, but the process for its development is generally applicable beyond EOCs and HRA. The BBN modeling framework allows capturing diverse influencing factor interactions, e.g., saturation and compensation of influences, going beyond the capabilities of the traditional multiplier-based models used in HRA.

Besides the actual BBN to support the EFI assessment, a contribution of the present work is the BBN development process itself. Two sets of operational event analyses represent the main empirical source for the BBN development. The sets provide patterns of factor ratings and corresponding error-forcing impact to be used to determine the BBN parameters. The BBN process develops the BBN around these examples, combining the functional interpolation method (a CPD filling algorithm) and expert judgment. The paper establishes a traceable process, in the use of the empirical basis and of the expert judgment. Traceability in all steps allows review of the process by other parties and, eventually, fosters the acceptance of the model. Traceability also allows the possibility to incorporate new evidence as it becomes available as well as different expert judgment assessments. The proposed process aligns well with the elements from [32] of third generation HRA methods, especially concerning the use of Bayesian parameter updating methods to formally incorporate additional and diverse novel evidence. Indeed, while the process is specified for operational event data, it can be directly extended to any form of data, which can be treated by a Bayesian update framework.

The two EOC sets have been populated by two different principal analysts and with large time separation (over 10 years between the two sets): fitting the BBN model to both sets makes its predictions less analyst-dependent and improves its generalization capability. The BBN development process is multi-stage. A first, Prior BBN is developed applying the interpolation method with input from expert judgment and evidence from EOC Set I related to the method anchors. The BBN is then updated to adjust to the whole EOC Set I (obtaining Posterior I BBN) and then to EOC Set II (obtaining Posterior II BBN). This allows intermediate validation (comparing the predictions of Posterior I BBN on EOC Set II) and puts forwards the steps for future BBN updates as new EOC events (or new analyst assessments) become available.

Having established a model to link influencing factor ratings and EFI, the next step for the EOC quantification model is to calibrate the EFI levels on failure probabilities. Future work will review the original CESA-Q calibration in view of new data becoming available through the on-going data collection activities [47–49].

CRediT authorship contribution statement

Luca Podofillini: Writing – original draft, Methodology, Conceptualization. **Bernhard Reer:** Writing – review & editing, Data curation, Conceptualization. **Vinh N. Dang:** Writing – review & editing, Supervision, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The work was funded by the Swiss Federal Nuclear Safety Inspectorate (ENSI), under contract Nr. 101163.

Disclaimer

The views expressed in this work are solely those of the authors.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ress.2022.108903.

Appendix A. CESA-Q: scaling guidance and EOC Sets I and II

See Table A1
See Table A2
See Table A3

Table A1
CESA-Q scaling guidance [38].

Factor[1,2]	Scores (levels)[2]	Situational features
VH	0 *(very strongly error-forcing)*	no hint to verify adequacy of motivated action; need to verify is clearly out of mind
	0.2 *(strongly error-forcing)*	unspecific or weak hint to verify under normal operating conditions: occasional checking practice (e.g., verify success of preceding action before proceeding with the next action, although this verification is not explicitly cued); or indication that something is deviant with rather weak association to required checking
		strongly degraded hint to verify; e.g., by urgency to act due to high time pressure
	0.5 *(moderately error-forcing)*	unspecific or weak hint to verify under emergency conditions, and verification subject (such as backup display of reactor pressure) relates to a safety parameter
	0.8 *(slightly error-forcing)*	viable but slightly degraded hint to verify; e.g., alarm with some ambiguity potential
	1 *(success-forcing)*	viable hint (e.g., clear instruction, alarm, feedback or warning tag) to verify adequacy of motivated action
VM	0 *(very strongly error-forcing)*	missing or essentially degraded (visibility, readability) indications of inadequacy of motivated action
	0.5 *(moderately error-forcing)*	degraded indications of inadequacy of motivated action: main indication unavailable, but backup indication available; or some visibility or readability problems
	1 *(success-forcing)*	main indication available and clearly visible
VD	N/A	relevant indications unavailable or major degradation of visibility
	0 *(very strongly error-forcing)*	implication of indications (which allow to identify inadequacy of motivated action) is totally unclear (e.g., due to major deficiency in training)
	0.5 *(moderately error-forcing)*	complex (knowledge-based) reasoning required to identify *implication*, e.g., due to the presence of conflicting information and poor or missing priority rules, or due to complexity of the rule on indication's implication, or due to masking events
	0.8 *(slightly error-forcing)*	cognitive requirement slightly increased; e.g., rule is clear but an unexpected parameter value leads to a deviation from the base case of trained rule application
	1 *(not error-forcing)*	implication of indications (of inadequacy of motivated action) is rather clear; identifiable from skill-based association, or written or recallable rule
VE	N/A	relevant indications unavailable or major degradation of visibility
	0 *(error-forcing)*	high physical effort (e.g., going to another location, or implementing specific valve alignments) required for verifying adequacy of motivated action
	1 *(not error-forcing)*	negligible physical effort required for verification (e.g., by referring to display in the vicinity)
TP	0 *(very strongly error-forcing)*	extremely high time pressure; decision time window (TW) in the order of seconds (typically) to deal with a potentially conflicting goal of viable concern
	0.2 *(strongly error-forcing)*	high time pressure; e.g., decision TW in the order of 1 min (typically) to deal with a potentially conflicting goal of viable concern
	0.5 *(moderately error-forcing)*	moderate time pressure; e.g., decision TW in the order of 10 min (typically) available to deal with a potentially conflicting goal of viable concern; or slight urgency to act in order to meet the outage schedule
	1 *(not error-forcing)*	no urgency to act; motivated action is not associated with a time constraint
BP	0 *(very strongly error-forcing)*	very high benefit clearly associated with motivated action because of management's expectations, e.g., practice for criticizing operators in case of economical loss due to rule compliance
	0.5 *(moderately error-forcing)*	high benefit clearly associated with motivated action because of major concurrent safety or operational goal; e.g., prospect of maintaining viable safety function, or prospect of prevention of major economical loss like plant shutdown for months
	1 *(not error-forcing)*	no particular benefit prospect
DP	0 *(not success-forcing)*	no particular damage potential in association with motivated action
	0.5 *(moderately success-forcing)*	immediate equipment damage potential of drastic (aversion-forcing) nature; e.g., fire
	1 *(very strongly success-forcing)*	immediate person injury potential of drastic (aversion-forcing) nature; e.g., high pressure steam escape close to place of work
PR	0 *(not success-forcing)*	no personal redundancy
		degraded personal redundancy; e.g., mainly restricted to formal checking requirement, or limited efficiency with respect to possible error modes covered by checking
	1 *(success-forcing)*	personal redundancy available (and nominally effective)

[1] Adjustment factor abbreviations: Verification hint (VH); Verification means (VM); Verification difficulty (VD); Verification effort (VE); Time pressure (TP); Benefit prospect (BP); Damage potential (DP); Personal redundancy (PR).
[2] This table maintains the original CESA-Q nomenclature of "Factors" and "Scores". In the BBN implementation, these become "Nodes" and "States".

Table A2
EOC cases Set I [38,39].

		CESA-Q Adjustment factors(2) and BBN intermediate factors(3), (4)											
Case ID(1)	**Event Title**	**VH**	**VM**	**VD**	**Ver Cog**	**PR**	**VE**	**Ver**	**BP**	**DP**	**BD**	**TP**	**EFI(5)**
AE.1	High Voltage Exposure	1	1	1	Very low	0	1	Very low	1	1	SF	1	Very low
AE.2	Fire and Loss of Offsite Power	1	1	1	Very low	1	0	High	1	0.5	MSF	1	High
AE.3	Loss of Coolant through Shutdown Cooling Suction Valve	0.2	1	1	Very high	1	1	Very high	1	0	Neutral	1	Very high
AE.4	Loss of Coolant through RCS Hot Leg	1	0.5	0.5	High	0	0	Very high	1	0	Neutral	1	Very high
AE.5	Loss of Coolant through RHR Discharge Isolation Valve	0	0.5	0.5	Ex. high	0	0	0	1	0	Neutral	1	Ex. high
AE.6	Person Injury and Loss of Coolant through Cleanup System	1	1	1	Very low	1	1	Very low	1	1	SF	1	Very low
AE.7	Loss of Coolant to Suppression Pool	0.2	0	0	Ex. high	0	0	Ex. high	1	0	Neutral	1	Ex. high
AE.8	Tar-Air Mixture Explosion	0	0	0	Ex. high	0	0	Ex. high	1	0	Neutral	1	Ex. high
AE.9	Reflux from Gas Purification to Demineralized Water System	1	1	0.5	High	1	1	High	1	0	Neutral	1	High
MI.1	Partial Boil-Off of Coolant Inventory	0.2	1	0.5	Very high	0	1	Very high	1	0	Neutral	1	Very high
MI.2	Loss of Coolant through Faulted Steam Generator	0.5	1	0.5	High	1	1	High	0.5	0	MEF	0.5	Very high
MI.3	Reactor Overheating due to Degradation of Safety Injection	0.8	1	0.5	High	1	1	High	1	0	Neutral	0.5	High
MI.4	Core Damage due to Termination of Safety Injection	0.2	0.5	0.5	Very high	1	1	Very high	0.5	0	MEF	0.2	Very high
AD.1	Vessel Over-Pressurization	0	1	1	High	1	1	High	1	0	Neutral	1	High
AD.2	Damage of High Pressure Injection Pumps	0.5	0.5	0.5	High	1	1	High	1	0	Neutral	0.5	High
AD.3	Degradation of Reactor Coolant Pump Integrity	0.5	1	0.5	High	1	1	High	1	0	Neutral	0.5	High
AD.4	Boron Dilution by Cleanup System Operation	0.2	1	1	High	1	0	High	1	0	Neutral	0.5	High
AD.5	Exothermic Release of Chemicals	0.5	1	0.5	High	1	1	High	1	0	Neutral	1	High

(1) The event coding reflects the dominant situational feature (MI: Misleading Indication or Instruction, AE: Adverse Exception, AD: Adverse Distraction).
(2) Coding for adjustment factors: VH: Verification Hints, VM: Verification Means, VD: Verification Difficulty, PR: Personal Redundancy, VE: Verification Effort, BP: Benefit Prospect, DP: Damage Potential, TP: Time pressure. Coding for scores in Table A2.
(3) In shaded cells, BBN nodes. VerCog: Verification (cognitive), Ver: Verification, BD: Benefit_Damage identify the BBN intermediate nodes added to group the adjustment factor effects. These factors do not appear in the original CESA-Q framework and have been introduced for the BBN model, see Section 3.1.
(4) See Fig. 2 for the definition of the states of the BBN intermediate nodes.
(5) Error Forcing Impact (EFI) levels are given in Table 1.

Table A3
EOC cases Set II [40].

		CESA-Q Adjustment factors(2) and BBN intermediate factors(3), (4)										
Case ID	Event title	VH	VM	VD	Ver Cog	PR	VE	Ver	BP	DP	BD	TP	EFI(5)
AE*.1	Cont. flood	0	1	1	Ex. high	1	1	Ex. high	1	0	Neutral	1	Ext High
AE*5	LOCA thru shutdown	0	1	1	Ex. high	1	1	Ex. high	1	0	Neutral	1	Ext High
MI*.1	Op. wrong trains	0.2	0	1	Very High	0	1	Very High	1	0	Neutral	1	Very High
MI*.5	Gadolinium inj.	0.2	0.5	1	Very High	0	1	Very High	1	0	Neutral	1	Very High
AE*.3	LOCA due to test	0.2	1	0	Ex. high	1	1	Ex. high	1	0	Neutral	1	Very High
AD*.1	LOCA thru bypass	0.2	1	0.5	Very High	1	1	Very High	0.2	0.5	Neutral	1	Ext High
AE*.2	Degrad. of neutron mon.	0.5	1	0.5	High	1	1	High	1	0	Neutral	0.5	High
MI*.2	Safety inj., faulty test	0.5	1	1	Low	0	0.8	High	1	0	Neutral	1	High
MI*.4	Drop of fuel	0.8	0.5	0.5	High	1	1	High	1	0.5	Neutral	1	Very High
AE*.4	Electr. lineup and fire	0.8	1	0.5	Low	1	1	Low	1	0	Neutral	1	Low
MI*.3	Safety inj., intlk.	0.8	1	1	Very low	0	1	Very low	1	0	Neutral	0.5	Low

(1) The event coding reflects the dominant situational feature (MI: Misleading Indication or Instruction, AE: Adverse Exception, AD: Adverse Distraction). The asterisk * is used to denote Set II events.
(2)Coding for adjustment factors: VH: Verification Hints, VM: Verification Means, VD: Verification Difficulty, PR: Personal Redundancy, VE: Verification Effort, BP: Benefit Prospect, DP: Damage Potential, TP: Time pressure. Coding for scores in Table A2.
(3)In shaded cells, BBN nodes. VerCog: Verification (Cognitive), Ver: Verification, BD: Benefit_Damage identify the BBN intermediate nodes added to group the adjustment factor effects. These factors do not appear in the original CESA-Q framework and have been introduced for the BBN model.
(4)See Fig. 2 for the definition of the states of the BBN intermediate nodes.
(5)Error Forcing Impact (EFI) levels are given in Table 1.

References

[1] Kirwan B. A guide to practical human reliability assessment. Boca Raton, FL, USA: CRC press; 1994.
[2] Spurgin AJ. Human reliability assessment – theory and practice. Boca Raton, FL, USA: CRC press; 2010.
[3] ASME. Standard for level 1 /large early release frequency probabilistic risk assessment for nuclear power plant applications. Standard of the American Society of Mechanical Engineers, ASME; 2008. RA-S - 2008(R2019)update 2018.
[4] Kolaczkowski A, Forester J, Lois E, Cooper S. Good practices for implementing human reliability analysis (HRA), NUREG-1792. Washington, DC: U.S. Nuclear Regulatory Commission; 2005.
[5] Forester J, Kolaczkowski A, Cooper S, Bley D, Lois E. ATHEANA user's guide, NUREG-1880. Washington, DC: U.S. Nuclear Regulatory Commission; 2007.
[6] EPRI/NRC EPRI/NRC-RES. Fire human reliability analysis guidelines. NUREG-1921. Washington, DC: U.S. Nuclear Regulatory Commission; 2012.
[7] Reer B, Dang VN, Hirschberg S. The CESA method and its application in a plant-specific pilot study on errors of commission. Reliab Eng Syst Saf 2004;83(2):187–205.
[8] Podofillini L, Dang VN, Nusbaumer O, Dres D. A pilot study for errors of commission for a boiling water reactor using the CESA method. Reliab Eng Syst Saf 2013;109:86–98.
[9] Reason J. Human error. UK: Cambridge University Press Cambridge; 1990. https://doi.org/10.1017/CBO9781139062367.
[10] Sträter O, Dang VN, Kaufer B, Daniels A. On the way to assess errors of commission. Reliab Eng Syst Saf 2004;83(2):129–38.
[11] Swain AD, Guttmann HE. Handbook of human reliability analysis with emphasis on. nuclear power plant applications. Washington, DC: U.S. Nuclear Regulatory Commission; 1983. NUREG/CR-1278.
[12] Technical basis and implementation guidelines for a technique for human event analysis (ATHEANA). Washington, DC, USA: U.S. Nuclear Regulatory Commission; 2000. NUREG-1624, Revision 1.
[13] Mosneron-Dupin F, Reer B, Heslinga G, Sträter O, Gerdes V, Saliou G, Ullwer W. Human-centred modeling in human reliability analysis: some trends based on case studies. Reliab Eng Syst Saf 1997;58:249–74.
[14] Gertman D, Blackman H, Byers J, Haney L, Smith C, Marble J. The SPAR-H method. Washington, DC: US Nuclear Regulatory Commission; 2005. NUREG/CR-6883.
[15] Williams JC. HEART – a proposed method for assessing and reducing human error. In: Proceedings of the 9th advances in reliability technology symposium. University of Bradford; 1986.
[16] Hollnagel E. Cognitive reliability and error analysis method (CREAM). Oxford, UK: Elsevier Science; 1998.
[17] Langseth H, Portinale L. Bayesian networks in reliability. Reliab Eng Syst Saf 2007;92(1):92–108.
[18] Fenton NE, Neil MD. Risk assessment and decision analysis with bayesian networks. Boca Raton, FL, USA: CRC Press; 2013.
[19] Kabir S, Papadopoulos Y. Applications of Bayesian networks and petri nets in safety, reliability, and risk assessments: a review. Saf Sci 2019;115:154–75. https://doi.org/10.1016/j.ssci.2019.02.009.
[20] Mkrtchyan L, Podofillini L, Dang VN. Bayesian belief networks for human reliability analysis: a review of applications and gaps. Reliab Eng Syst Saf 2015;139:1–16.
[21] Zwirglmaier K, Straub D, Groth KM. Capturing cognitive causal paths in human reliability analysis with Bayesian network models. Reliab Eng Syst Saf 2017;158:117–29. https://doi.org/10.1016/j.ress.2016.10.010.
[22] Groth K, Mosleh A. Deriving causal Bayesian networks from human reliability analysis data: a methodology and example model. J Risk Reliab 2012;226:361–79.
[23] Sundaramurthi R, Smidts C. Human reliability modeling for the next generation system code. Ann Nucl Energy 2013;52:137–56.
[24] Zhang X, Mahadevan S. Bayesian network modeling of accident investigation reports for aviation safety assessment. Reliab Eng Syst Saf 2021;209:107371.
[25] Zhang G, Thai VV, Law AWK, Yuen KF, Loh HS, Zhou Q. Quantitative risk assessment of seafarers' nonfatal injuries due to occupational accidents based on bayesian network modeling. Risk Anal 2020;40(1):8–23.
[26] Simsekler MCE, Qazi A. Adoption of a data-driven Bayesian belief network investigating organizational factors that influence patient safety. Risk Anal 2022;42(6):1277–93. https://doi.org/10.1111/risa.13610.
[27] Steijn WMP, van Kampen JN, van der Beek D, Groeneweg J, van Gelder PHAJM. An integration of human factors into quantitative risk analysis using Bayesian belief networks towards developing a 'QRA+'. Saf Sci 2020;122:104514. https://doi.org/10.1016/j.ssci.2019.104514.
[28] Asadayoobi N, Taghipour S, Jaber MY. Predicting human reliability based on probabilistic mission completion time using Bayesian Network. Reliab Eng Syst Saf 2022;221:108324.
[29] Rosário CR, Gonçalves Amaral F, Malmann Kuffel FJ, Mahlmann Kipper L, Frozza R. Using Bayesian belief networks to improve distributed situation

awareness in shift changeovers: a case study. Expert Syst Appl 2022;188:116039. https://doi.org/10.1016/j.eswa.2021.116039. ISSN 0957-4174.
[30] Wang Z, Zeng S, Guo J, Che H. A Bayesian network for reliability assessment of man-machine phased-mission system considering the phase dependencies of human cognitive error. Reliab Eng Syst Saf 2021;207:107385.
[31] Abrishami S, Khakzad N, Hosseini SM, van Gelder PHAJM. BN-SLIM: a Bayesian network methodology for human reliability assessment based on success likelihood index method (SLIM). Reliab Eng Syst Saf 2020;193:106647.
[32] Groth KM, Smith R, Moradi R. A hybrid algorithm for developing third generation HRA methods using simulator data, causal models, and cognitive science. Reliab Eng Syst Saf 2019;191:106507. https://doi.org/10.1016/j.ress.2019.106507.
[33] Barons MJ, Mascaro S, Hanea AM. Balancing the elicitation burden and the richness of expert input when quantifying discrete Bayesian networks. Risk Anal 2021;42(6):1196–234. https://doi.org/10.1111/risa.13772.
[34] Werner C, Bedford T, Cooke R, Hanea AM, Morales-Napoles O. Expert judgement for dependence in probabilistic modelling: a systematic literature review and future research directions. Eur J Oper Res 2016;258(3):801–19.
[35] Mkrtchyan L, Podofillini L, Dang VN. Methods for building conditional probability tables of Bayesian belief networks from limited judgment: an evaluation for human reliability application. Reliab Eng Syst Saf 2016;151:93–112. https://doi.org/10.1016/j.ress.2016.01.004.
[36] Cooke RM. Experts in uncertainty. New York: Oxford university press; 1991.
[37] Pearl J. Fusion, propagation, and structuring in belief networks. Artif Intell 1986;29(3):241–88.
[38] Reer B. Outline of a method for quantifying errors of commission, lea 09-302. Villigen PSI, Switzerland: Paul Scherrer Institut; 2009.
[39] Reer B, Dang VN. Situational features of errors of commission identified from operating experience. lea 09-303. Villigen PSI, Switzerland, Villigen PSI, Switzerland: Paul Scherrer Institut; 2009.
[40] Podofillini L, Reer B, Dang VN. Analysis of recent operational events involving inappropriate actions: influencing factors and root causes. Reliab Eng Syst Saf 2021;216:108013. https://doi.org/10.1016/j.ress.2021.108013.
[41] Mosleh A. Bayesian modeling of expert-to-expert variability and dependence in estimating rare event frequencies. Reliab Eng Syst Saf 1992;38:47–57.
[42] Podofillini L, Dang VN. A Bayesian approach to treat expert-elicited probabilities in human reliability analysis model construction. Reliab Eng Syst Saf 2013;117:52–64.
[43] Groth KM, Smith CL, Swiler LP. A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods. Reliab Eng Syst Saf 2014;128:32–40. https://doi.org/10.1016/j.ress.2014.03.010.
[44] Agena Ltd, AgenaRisk Software. www.agenarisk.com. 2007.
[45] Parry G.W., Lydell B.O. An approach to the analysis of operator actions in probabilistic risk assessment. EPRI TR-100259, Palo Alto, CA, USA. 1992.
[46] Xing J, Parry G, Presley M, Forester J, Hendrickson S, Dang V. An integrated human event analysis system (IDHEAS) for nuclear power plant internal events at-power application (Volume 1). NUREG-2199. Washington, DC, USA: U.S. Nuclear Regulatory Commission; 2017.
[47] Chang JY, Bley D, Criscione L, Kirwan B, Mosleh A, Madary T, Nowell R, Richards R, Roth ER, Sieben S, Zoulis A. The SACADA database for human reliability and human performance. Reliab Eng Syst Saf 2014;125:117–33.
[48] Jung W, Park J, Kim J, Choi SY, Kim S. HuREX – a framework of HRA data collection from simulators in nuclear power plants. Reliab Eng Syst Saf 2020;194:106235.
[49] Kim Y, Choi SY, Park J, Kim J. Empirical study on human error probability of procedure-extraneous behaviors. Reliab Eng Syst Saf 2022;227:108727.